Lessons Learned for Improving Accessibility in Assessments for Students with Disabilities

U.S. Department of Education

In partnership with

Council of Chief State School Officers

May 22, 2012 • Washington, DC

Please silence all cell phones and pagers.

Thank you!

Introduction and Meeting Goals

***Alexa Posny**, Assistant Secretary*

Office of Special Education and Rehabilitative Services

***Ann Whalen**, Director of Policy and Program Implementation*

Implementation and Support Unit

Introduction

- Today's meeting is held in partnership:
  - Office of Special Education and Rehabilitative Services (OSERS)
  - Implementation and Support Unit (ISU)
  - Council of Chief State School Officers' Assessing Special Education Services (ASES) State Collaborative on Assessments and Student Standards (SCASS)

- Funded in part by The William and Flora Hewlett Foundation

Goals for the Meeting

- Share lessons learned from research and assessment development under previous grants
- Offer assistance to consortia developing next-generation assessment systems
- Format:
  - Five former grantees presenting research and lessons learned on improving accessibility of general assessments for students with disabilities
  - Table discussion with representatives from the consortia: PARCC, Smarter Balanced, the National Center and State Collaborative, and Dynamic Learning Maps

Consortia

- Awarded in September 2010
- Four-year grants to develop next-generation assessment systems in English language arts and mathematics

General Supervision Enhancement Grants

- Assessments for students with the most significant cognitive disabilities
- Two consortia, comprising 31 states and DC
  - National Center and State Collaborative (NCSC)
  - Dynamic Learning Maps

Race to the Top Assessment

- Two consortia representing 45 states and DC
  - Partnership for Assessment of Readiness of College and Careers (PARCC)
  - Smarter Balanced Assessment Consortium (Smarter Balanced)

IDEA

Who Are Children with Disabilities?

IDEA provides services for 6.9 million individuals with disabilities from birth through age 21

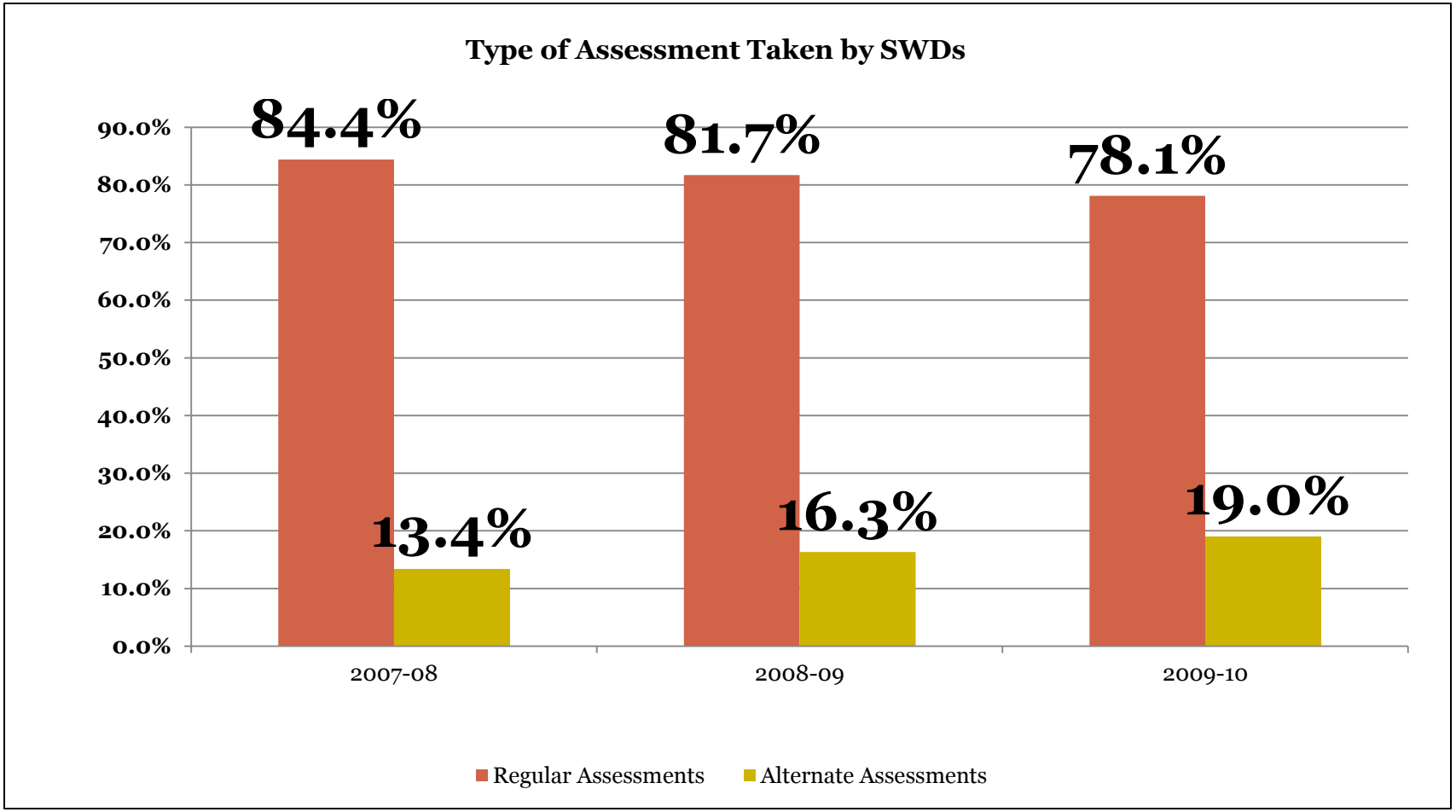
- 3% of population birth to age 2
- 14% of public school enrollees

13 Disability Categories

- 41% Specific Learning Disabilities
- 19% Speech or Language Impairments
- 17% All other categories
- 8% Intellectual Disabilities
- 7% Emotional Disturbance
- 6% Autism
- 2% Multiple Disabilities

Sources: United States Department of Education, Office of Special Education Programs, Individuals with Disabilities Education Act Data Accountability Center, 2010. U.S. National Center for Education Statistics, Digest of Education Statistics, annual, and Projections of Education Statistics, 2010.

How are SWDs participating in Statewide Assessments?



Source: United States Department of Education, Office of Special Education Programs, Individuals with Disabilities Education Act Data Accountability Center, 2008-2010.

Accountability:

Holding All Students to High Standards

SWDs can meet college- and career-ready standards:

- SWDs can excel within the general curriculum.
- SWDs can be prepared for success in post-secondary including college and/or careers.

SWDs must be ensured:

- Access to the general curriculum
- Ability to meet their unique needs
- IEPs with goals aligned to grade-level academic standards
- Teachers and support personnel able to deliver high-quality, evidence-based, individualized instructional and support services

Sources: Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools (NCEE 2009-4060). Wanzek, J. & Vaughn, S. (2010). Tier 3 interventions for students with significant reading problems. *Theory Into Practice, 49*, 305-314. Common Core State Standards Initiative. Application to Students with Disabilities. http://www.corestandards.org

Student Outcome Increases (1998-2009)

	All Students	Students with Disabilities
4th grade reading	8%	8%
8th grade math	10%	6%
Graduation rates	3.4%	10%
Post secondary enrollment	8%	38%
4 year college enrollment	14%	13%

RTTA Public Meetings

- This is the fourth RTTA public meeting.
  - April 15– Technology infrastructure
  - June 10– Automated scoring
  - August 10– Accessibility for students with disabilities and English learners
- The purpose of the meetings is to:
  - To support collaboration and provide technical assistance to PARCC and Smarter Balanced as they develop new assessment systems
  - Expand the knowledge and expertise of the Department and the public around key assessment issues
  - Facilitate discussion of key components of the systems with experts

About RTTA

- Support states in delivering a system of more effective and instructionally useful assessments that:
  - Provide accurate information about what students know and can do by:
    - Eliciting complex student demonstrations or applications of knowledge and skills, as appropriate
    - Accurately measuring student achievement across the full performance continuum
    - Accurately measuring student growth over a full academic year or course;
    - Helping educators determine whether individual students are ready for college and careers by the time of high school graduation and, in previous grade levels, whether they are on-track for readiness
  - Reflect good instructional practice and support a culture of continuous improvement
  - Effectively assess all students, including students with disabilities and English learners

Looking Forward

- Assessment systems must include one or more summative assessment components that are fully implemented by every state in each consortium by SY 2014-15, and are administered at least once during the academic year in, at a minimum:
  - English language arts and mathematics
  - Grades 3-8 and high school
- Results used to inform:
  - Teaching, learning, and program improvement
  - Determinations of school effectiveness
  - Determinations of principal and teacher effectiveness for the purposes of evaluation and support
  - Determinations of individual student college- and career-readiness

Students with Disabilities in RTTA

- The absolute priority requires that consortia create assessments for all students, including students with disabilities and English learners
- The consortia are required to develop tests accessible for these populations and to create and standardize accommodations policies
- PARCC and Smarter Balanced have established panels of external experts knowledgeable about the needs of students with disabilities to inform assessment development

Meeting Agenda

8:30-9:00	Welcome/setting the stage
9:00-9:30	Inclusive assessment: Considerations
9:30-10:00	Universal Design for Learning
10:00-10:15	Break
10:15-10:45	Cognitive labs and Opportunity-to-Learn studies
10:45-11:15	Innovative Items
11:15-11:45	Considerations in Assessing Low-Performing Students with Disabilities
11:45-12:00	Concluding Comments & Wrap-up
12:00	Adjourn

Invited Experts

- Lou Danielson, American Institutes for Research
- Stephen Elliott, Arizona State University
- Steve Ferrara, Pearson
- Sheryl Lazarus, National Center for Educational Outcomes at the University of Minnesota
- Shelley Loving-Ryder, Virginia Department of Education

Reminders

- Please place all cell phones and other devices on vibrate
- Resources:
  - Race to the Top Assessment: www2.ed.gov/programs/racetothetop-assessment
  - NCSC: www.cehd.umn.edu/nceo/projects/NCSC/NCSC.html
  - Dynamic Learning Maps: www.dynamiclearningmaps.org
- The purpose of this event is to promote a full discussion and hear a wide range of viewpoints on creating valid, reliable, and fair assessments for English learners and students with disabilities, as well as the challenges and opportunities afforded by the Race to the Top Assessment program. Through this meeting, the U.S. Department of Education is not seeking to promote and/or endorse any particular program, project, methodology or approach to this work.

BREAK

Thank you!!

Alexa Posny

U.S. Department of Education

Reminders

- Transcript and presentations from today's meeting will be available at:

www2.ed.gov/programs/racetothetop-assessment

- Written input may be submitted to racetothetop.assessment@ed.gov

Future Public Meetings

- Future meetings may focus on:
  - Interoperability and technology standards
  - Selection of a uniform growth model consistent with test purpose, structure, and intended uses
  - Setting achievement standards and performance level descriptors
- Information about future meetings will be posted on ed.gov and shared with stakeholder groups and prior meeting participants

AMERICAN INSTITUTES FOR RESEARCH®

Ensuring General Assessment Access for Students with Disabilities: Overview and Considerations

Louis Danielson, Ph.D.
Managing Director

May 2012

Copyright © 2011
American Institutes
for Research.
All rights reserved.

Timeline

1975	1991	1996	1997	2001
P.L. 94-142 enacted	NCEO funded	NAEP establishes students with disabilities accommodations	IDEA mandates assessments	NCLB passed by Congress

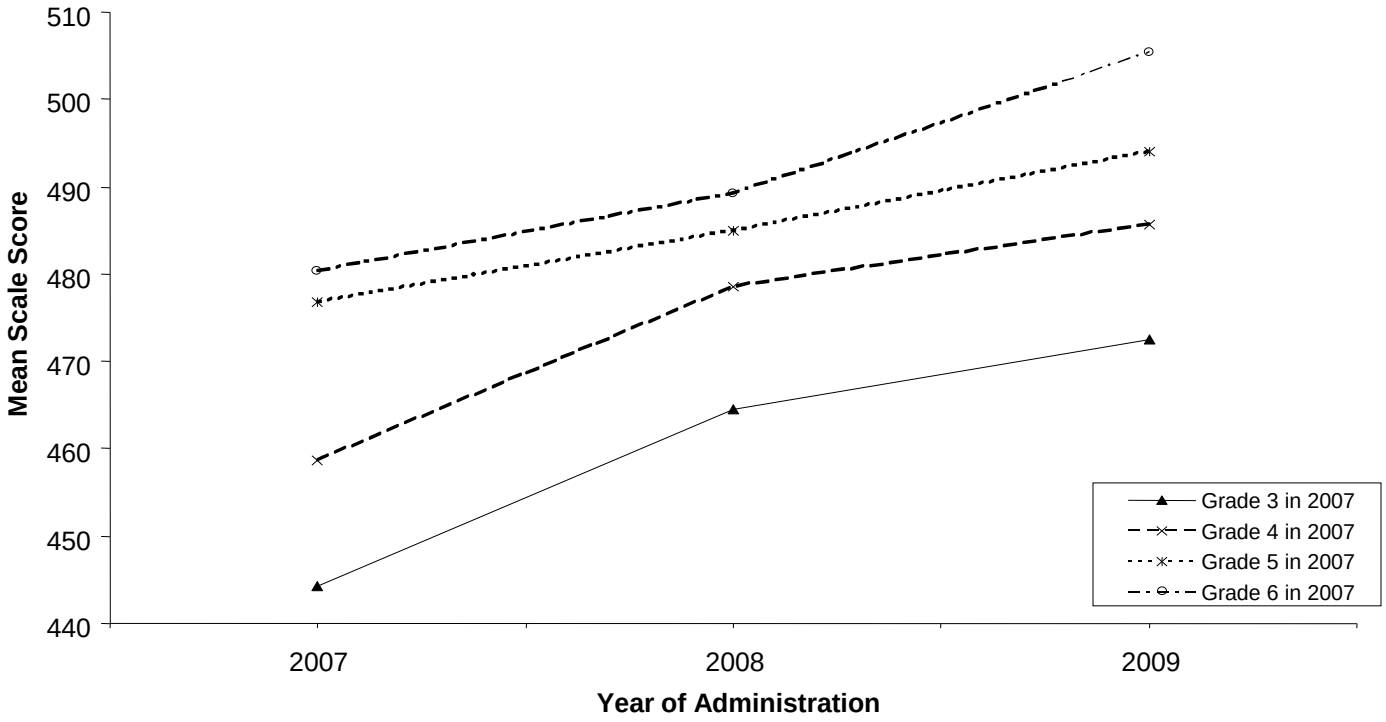
Key
IDEA=Individuals with Disabilities Education Act NAEP=National Assessment of Educational Progress NCLB=No Child Left Behind Act NCEO=National Center on Educational Outcomes P. L. 94-142=Education of All Handicapped Children Act (now known as IDEA)

Advances in Assessment of Students With Disabilities

- Alternative assessments
- Accommodations and universal design principles(UDL)
- Adaptive assessment
- Participation

Growth on the Vertical Scale: New Mexico Mathematics

New Mexico Alternative Assessment Mathematics Longitudinal Growth, by Grade



Continuing Issues in Assessment for SWD

- Continuous progress monitoring vs. one time assessment
- Portfolios for alternative assessment
- Consistent accommodation policies across states
- Out-of-level testing
- Assessing higher order content

Issues in Accountability Related to Students With Disabilities

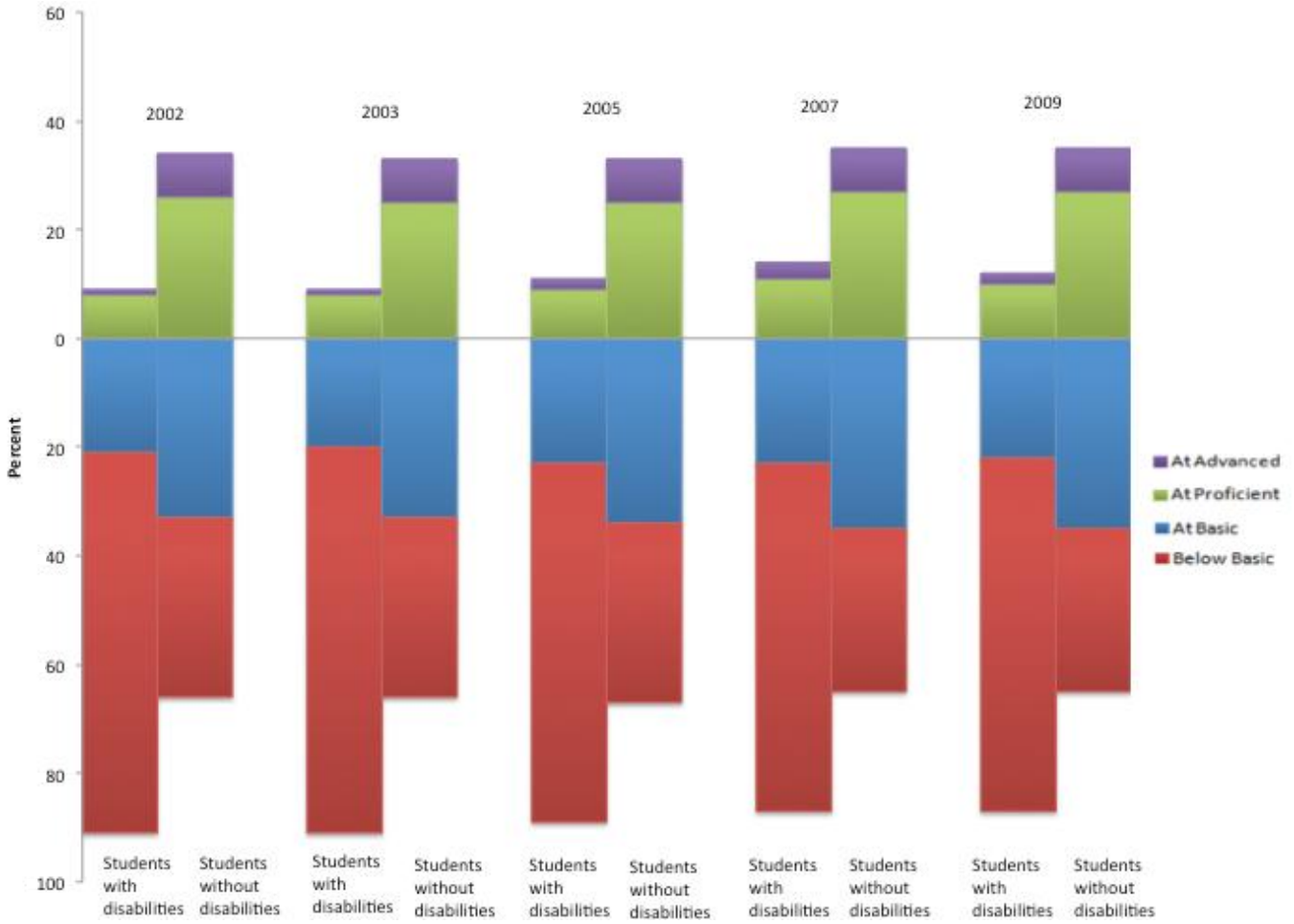
- “n” size
- Standards setting
- Fixed vs. progress standards
- Challenging expectations

The Bottom Line

Improve performance of students with disabilities

- Maintain challenging expectations for each SWD
- Create accountability system that incentivizes improvements in instruction

Reading Achievement Levels at Grade 4 by Student Disability Status: Various Years, 2002–2009



Note: Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002, 2003, 2005, 2007 and 2009 Reading Assessments.

Looking Ahead

- Developing improved assessments
- Developing alternative forms of accountability
- Improving instruction

9

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Louis Danielson, Ph.D.

P: 202-403-5850

E-Mail: ldanielson@air.org

1000 Thomas Jefferson Street, NW

Washington, DC 20007-3835

General Information: 202-403-5000

Website: www.air.org

Incorporating Universal Design Principles in Next Generation Assessment Item Design

Sheryl Lazarus

National Center on Educational Outcomes

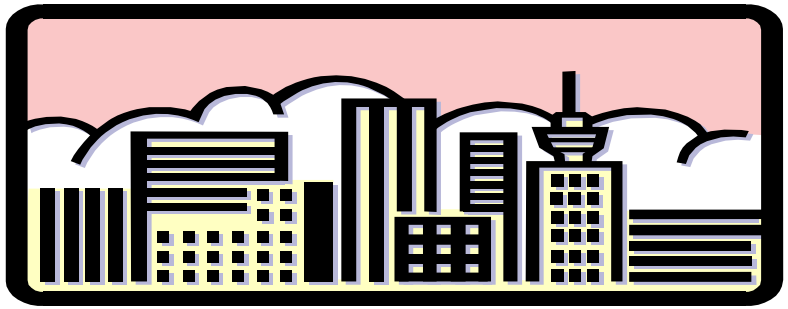
Public Meeting:
Lessons Learned for Improving Access to General Assessments by Low-Performing Students with Disabilities

Washington DC
May 22, 2012

Universal Design

The design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design.

Center for Universal Design (1997)



Think about universal design in architecture and tool design

- Curb cuts and ramps
- Elevators that talk to you
- Door handles rather than knobs
- Special pen shapes that are easier to hold

Universal Design for Learning (UDL)

- A set of principles for curriculum development that give all individuals equal opportunity to learn
- Provides a blueprint for creating instructional goals, methods, materials, and assessments that work for everyone.

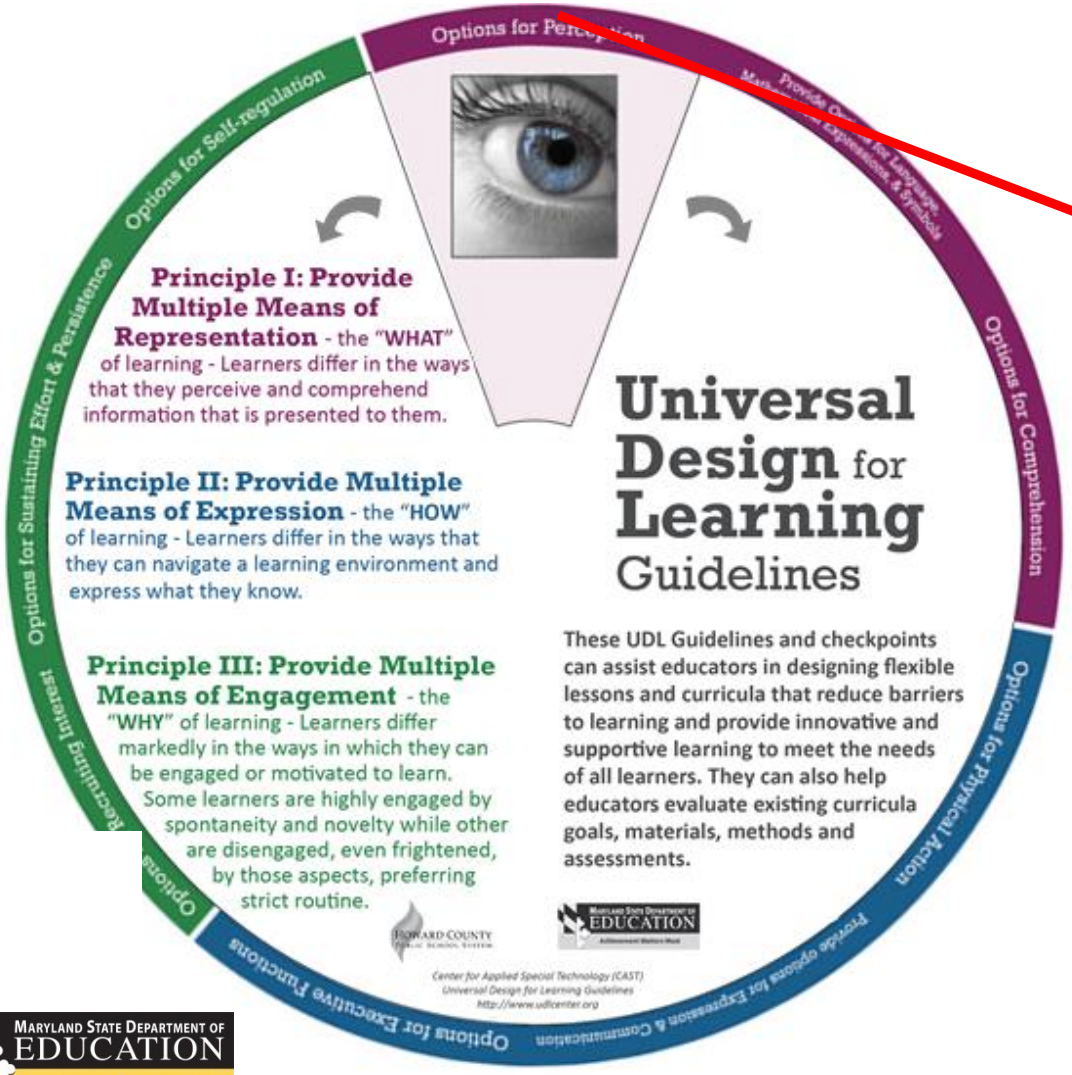
National Center on Universal Design for Learning

Components of UDL

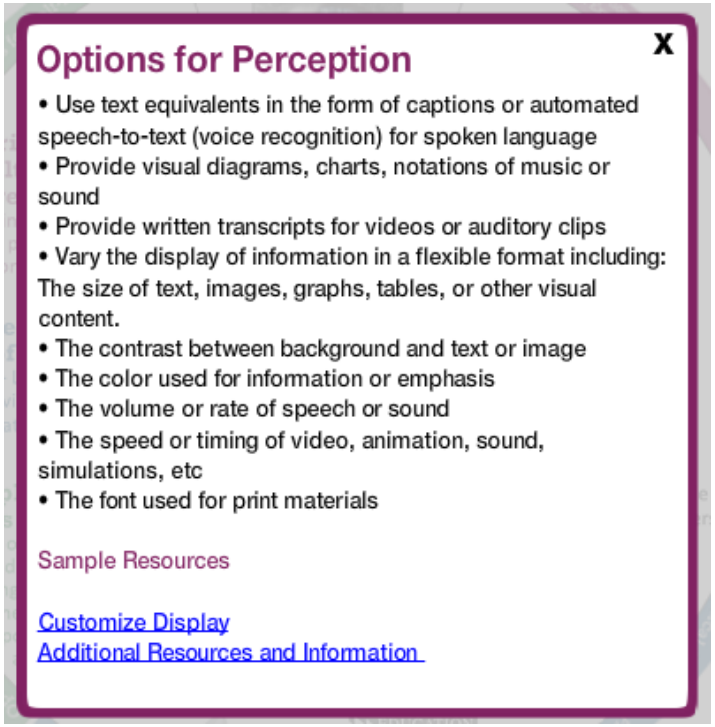
- Goals—learning expectations. They represent the knowledge, concepts, and skills all students should master, and are generally aligned to standards.
- Methods—instructional decisions, approaches, procedures that expert teachers use to accelerate or enhance learning.
- Materials—media used to present learning content
- Assessment—process of gathering information about a learner's performance using a variety of methods.

National Center on Universal Design for Learning

Universal Design for Learning: Tool for Teachers



Includes sample strategies for classrooms



MARYLAND STATE DEPARTMENT OF EDUCATION
Preparing World-Class Students

http://udlwheel.mdonlinegrants.org/

Universally designed assessments:

- are *designed from the beginning* to be accessible and valid for the widest range of students
- provide optimal standard assessment conditions

Universally Designed Assessments (UDA)

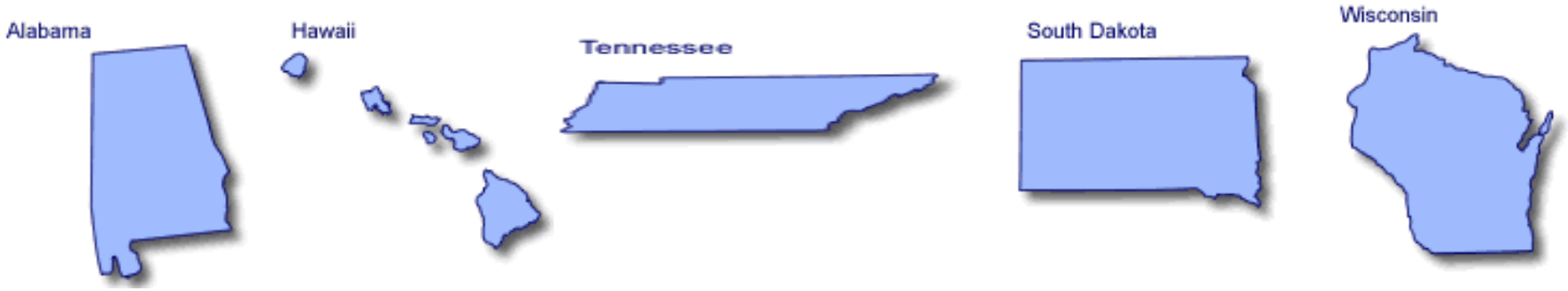
Tests that remain true to constructs, are easy to understand, and contain language that is accessible to all will give the truest readings of what students do and do not know. Universal Design does not mean "dumbing down" a test.

Who Benefits?

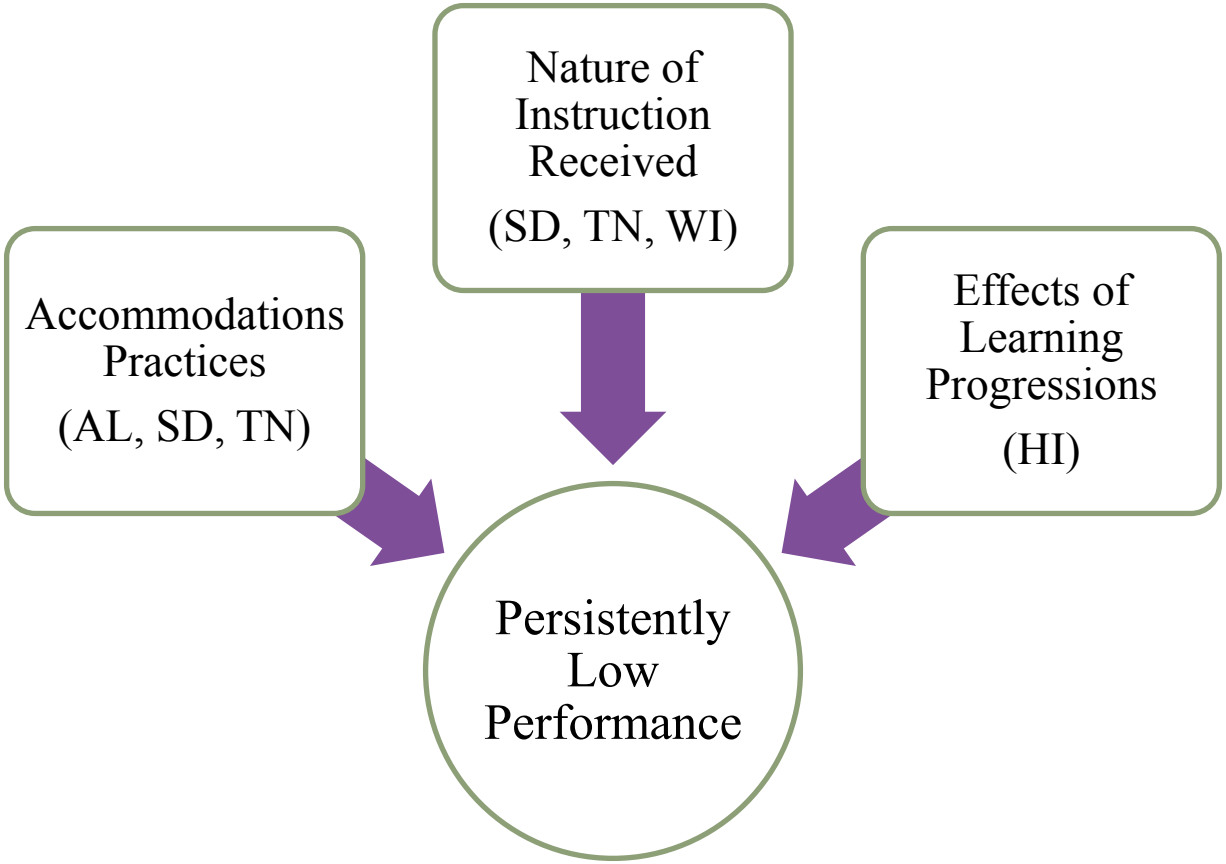
- Universal design does not apply exclusively to people with disabilities or limited English proficiency
- It applies to all individuals, with wide ranging characteristics

Multi- State GSEG Consortium

NATIONAL CENTER ON EDUCATIONAL OUTCOMES



Multi- State GSEG Consortium



Alabama GSEG

- What strategies can improve the universal design of assessments for low performing students with disabilities?
- What is universal design? . . . vs. . . . What is a lowered achievement expectation?

Elements of UD Assessments

- Inclusive assessment population
- Precisely defined constructs
- Accessible, non-biased items
- Amenable to accommodations

Elements of UD Assessments (continued)

- **Simple, clear, and intuitive instructions and procedures**
- **Maximum readability and comprehensibility**
- **Maximum legibility**

Thompson, Johnstone & Thurlow (2002)

UD Assessment Items

- Measures what it intends to measure
- Respect diversity of assessment population
- Have clear format for text
- Have clear pictures and graphics (when essential to item)
- Have concise and readable text
- Allow changes to its format without changing its meaning or difficulty

Thompson, Johnstone, Anderson, & Miller (2005)

Additional UDA Considerations for Computer-based Tests

- Layout and design
- Navigation
- Computer capabilities

Recommendations

- Incorporate elements of UD in early stages of development
- Include disability and language acquisition experts in items reviews
- Provide professional development for item developers and reviewers on UD
- Present the items being reviewed in format they will appear on test
- Include standards being tested with the items being reviewed
- Try out items with students
- Field test items in accommodated formats

Thompson, Johnstone, Anderson, & Miller (2005)

The Opportunity

- UD can help ensure that assessments do not restrict learning opportunities.
- UD can help us rethink how to create assessment items that provide a more accurate picture of what students know and can do.

For More Information

National Center on Educational Outcomes

www.nceo.info

Sheryl Lazarus

laza0019@umn.edu

References

- Center for Universal Design (1997). What is universal design? Center for Universal Design, North Carolina State University.

- Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments* (Technical Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

IMPROVING ACCESSIBILITY TO LARGE SCALE ASSESSMENTS FOR ALL STUDENTS

Stephen N. Elliott

Mickelson Foundation Professor & Director
Learning Sciences Institute
Arizona State University

May 2012

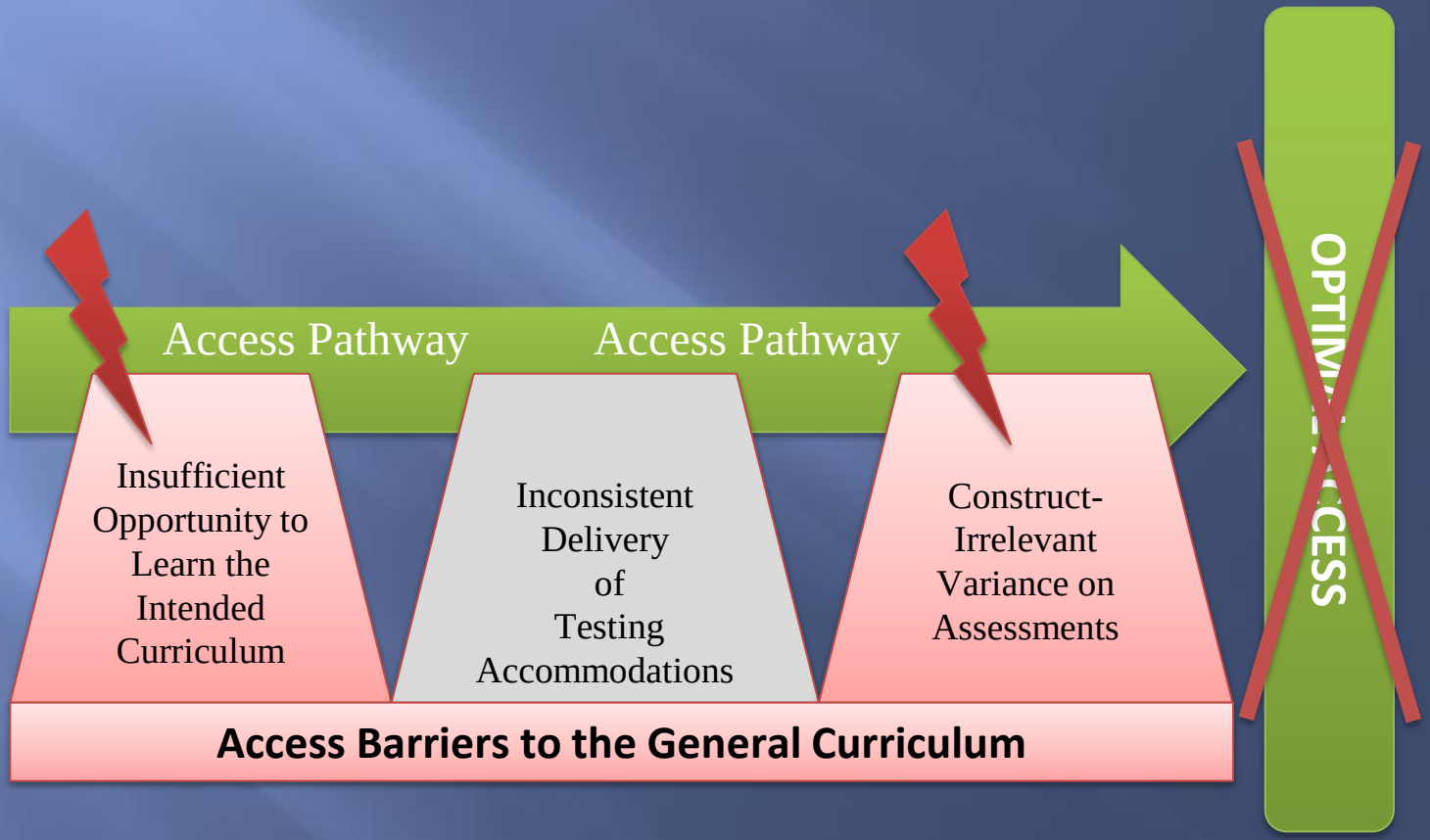
1. How can tests be modified to improve the measurement of knowledge and skills of students with persistent academic difficulties and who would qualify to take an AA-MAS, if available?

2. How can testing conditions be improved to increase the likelihood that all students have had the opportunity to learn the content standards that the tests are designed to measure?

Projects & Partners

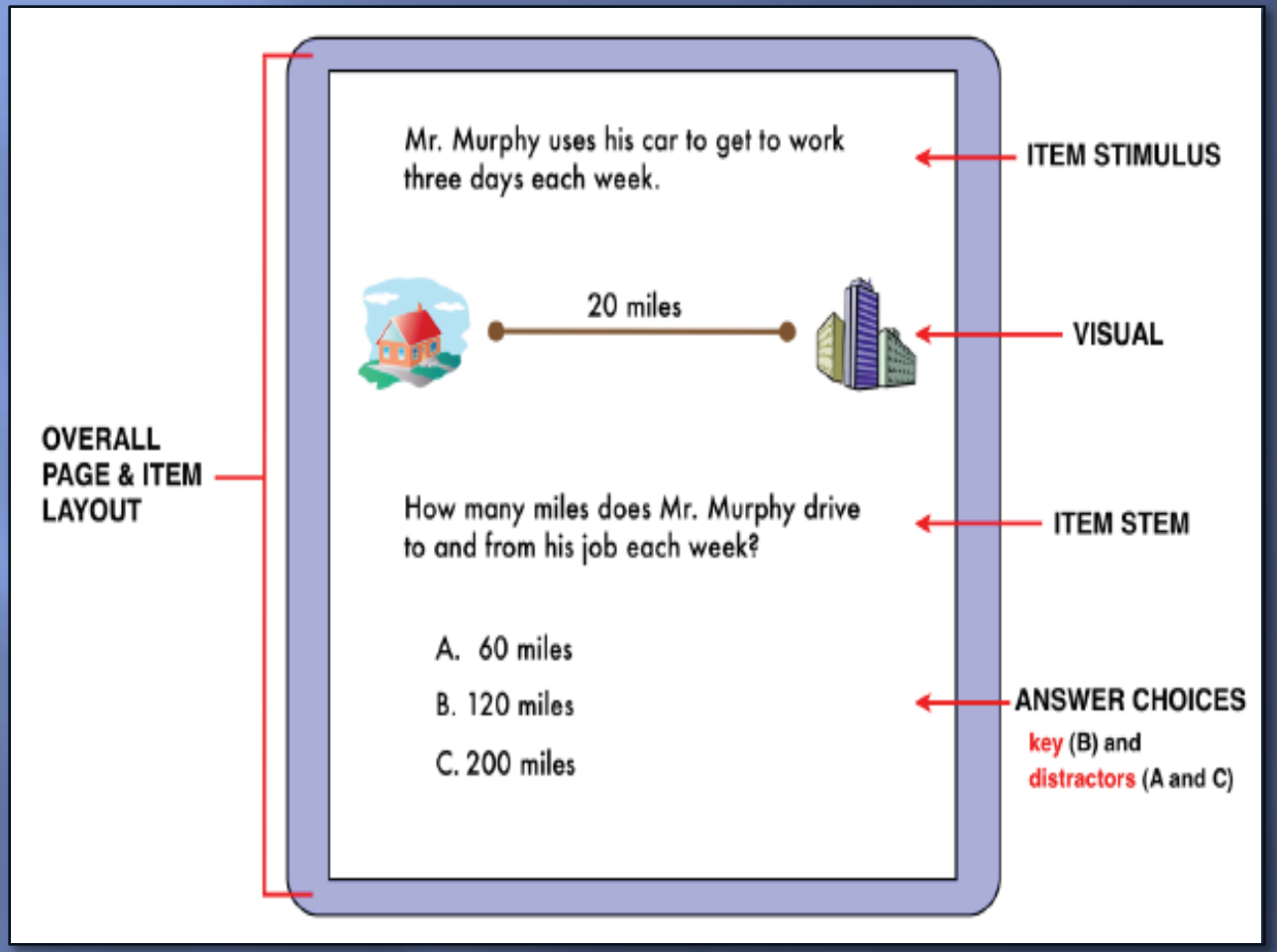
- **CAAVES: Consortium for Alternate Assessment Validity and Experimental Studies**
  - USDE funded; 2006-2009 [Award #S368A060012]
  - Partners: AZ, HI, ID, & IN + Vanderbilt Measurement Group + Discovery Education Assessment
- **CMAADI: Consortium for Modified Alternate Assessment Development and Implementation**
  - USDE funded; 2007-2011 [Award #H373X070026]
  - Arizona Dept. of Education
  - Indiana Dept. of Education
- **MAAPS: Modified Alternate Assessment Performance Screening project**
  - USDE funded; 2009-2011 [Award #S368A090006]
  - University of Pittsburgh
  - Discovery Education Assessment
  - Arizona Dept. of Education
  - Pennsylvania Dept. of Education
  - South Carolina Dept. of Education

Barriers to Access



* Adapted from Kurz & Elliott (2011)

Anatomy of an Item



Guiding Theories & Research for Item Modifications

- Evidenced-based model of test score validity,
- Universal design principles,
- Cognitive Load Theory for designing instructional materials, and
- Item writing research and practices.

Characteristics of Appropriate Modifications

Design Elements

- Simplify words and text structure
- Delete extraneous words
- Improve visuals and locate within item
- Use bold text for important words
- Examine answer distractors for plausibility and disproportionality of selection

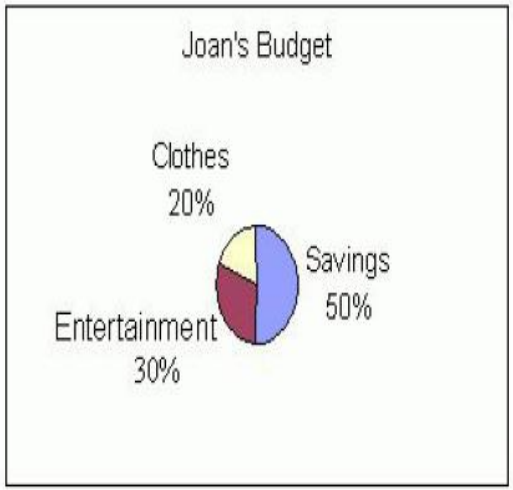
Desired Outcomes

- Increase accessibility
- Decrease item difficulty
- Increase item discrimination
- Increase reliability estimates
- Reduce readability level w/in grade range
- Maintain alignment w/ content stds.
- Maintain DOK for all items
- Reduce need for accommodations
- Reduce number of words; Improve students' perceptions of tests & motivation to engage in testing

The result: Increased test score validity!

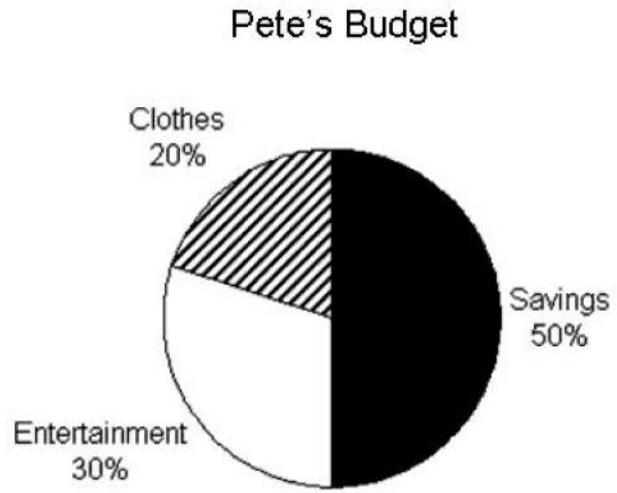
Example: Original to Modified Item

4. Joan earns $100 per month working part-time in a music store. Look at the pie chart that shows how Joan budgets her money each month. If Joan sticks to her budget, how much can she spend on clothes and entertainment each month?



- A. $50.00
- B. $40.00
- C. $30.00
- D. $20.00

4. Pete earns $100 per month. Based on the graph, how much can he spend on clothes and entertainment each month?

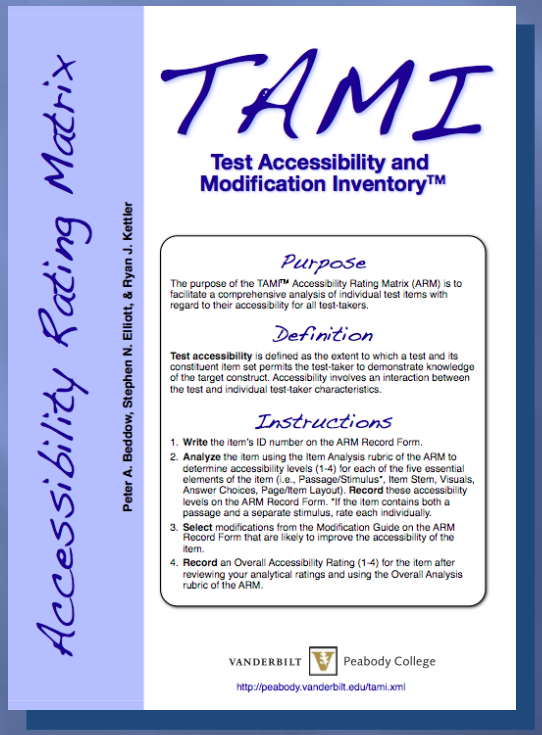


- A. $50.00
- B. $30.00
- C. $20.00

TAMI-ARM

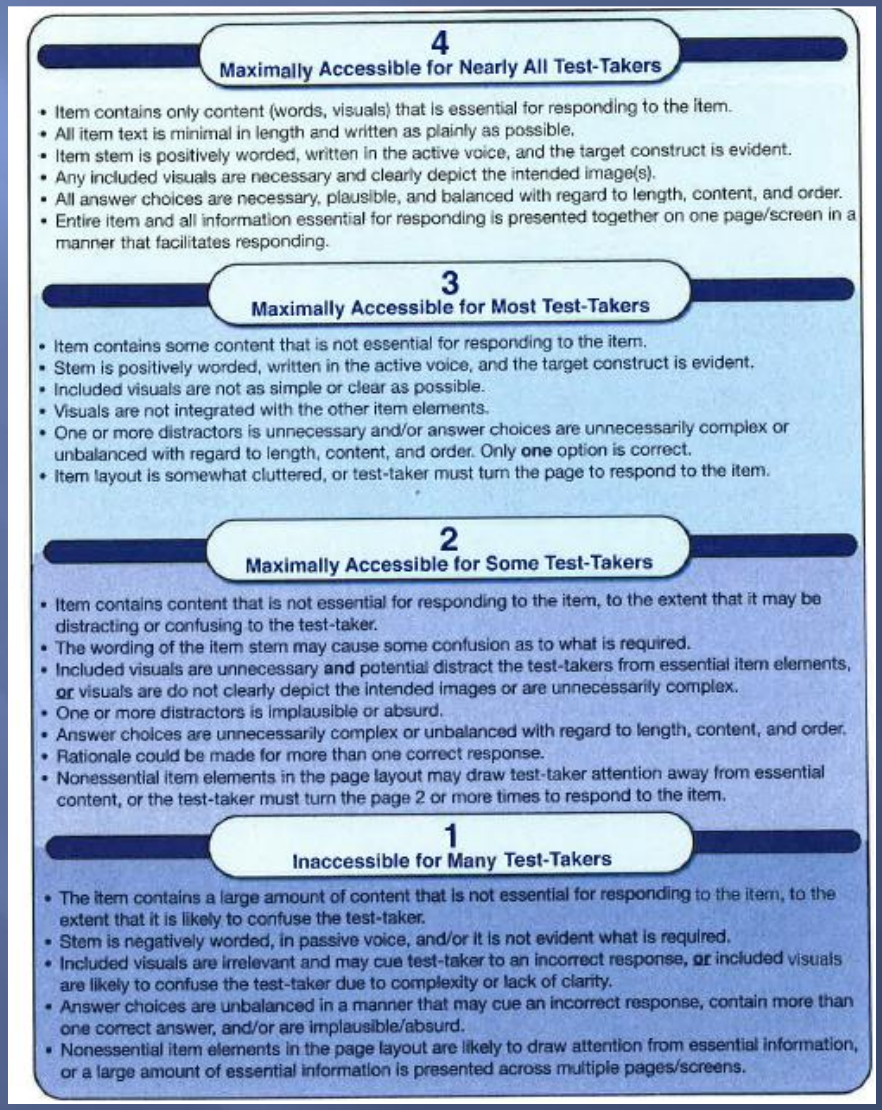
Test Accessibility and Modification Inventory-Accessibility Rating Matrix™

Test accessibility is defined as the extent to which a test and its constituent item set permits the test taker to demonstrate knowledge of the target construct. Accessibility involves an interaction between the test and individual test-taker characteristics.



Peabody.vanderbilt.edu/tami or search "**Vanderbilt TAMI**"

TAMI's Accessibility Levels



4
Maximally Accessible for Nearly All Test-Takers

- Item contains only content (words, visuals) that is essential for responding to the item.
- All item text is minimal in length and written as plainly as possible.
- Item stem is positively worded, written in the active voice, and the target construct is evident.
- Any included visuals are necessary and clearly depict the intended image(s).
- All answer choices are necessary, plausible, and balanced with regard to length, content, and order.
- Entire item and all information essential for responding is presented together on one page/screen in a manner that facilitates responding.

3
Maximally Accessible for Most Test-Takers

- Item contains some content that is not essential for responding to the item.
- Stem is positively worded, written in the active voice, and the target construct is evident.
- Included visuals are not as simple or clear as possible.
- Visuals are not integrated with the other item elements.
- One or more distractors is unnecessary and/or answer choices are unnecessarily complex or unbalanced with regard to length, content, and order. Only **one** option is correct.
- Item layout is somewhat cluttered, or test-taker must turn the page to respond to the item.

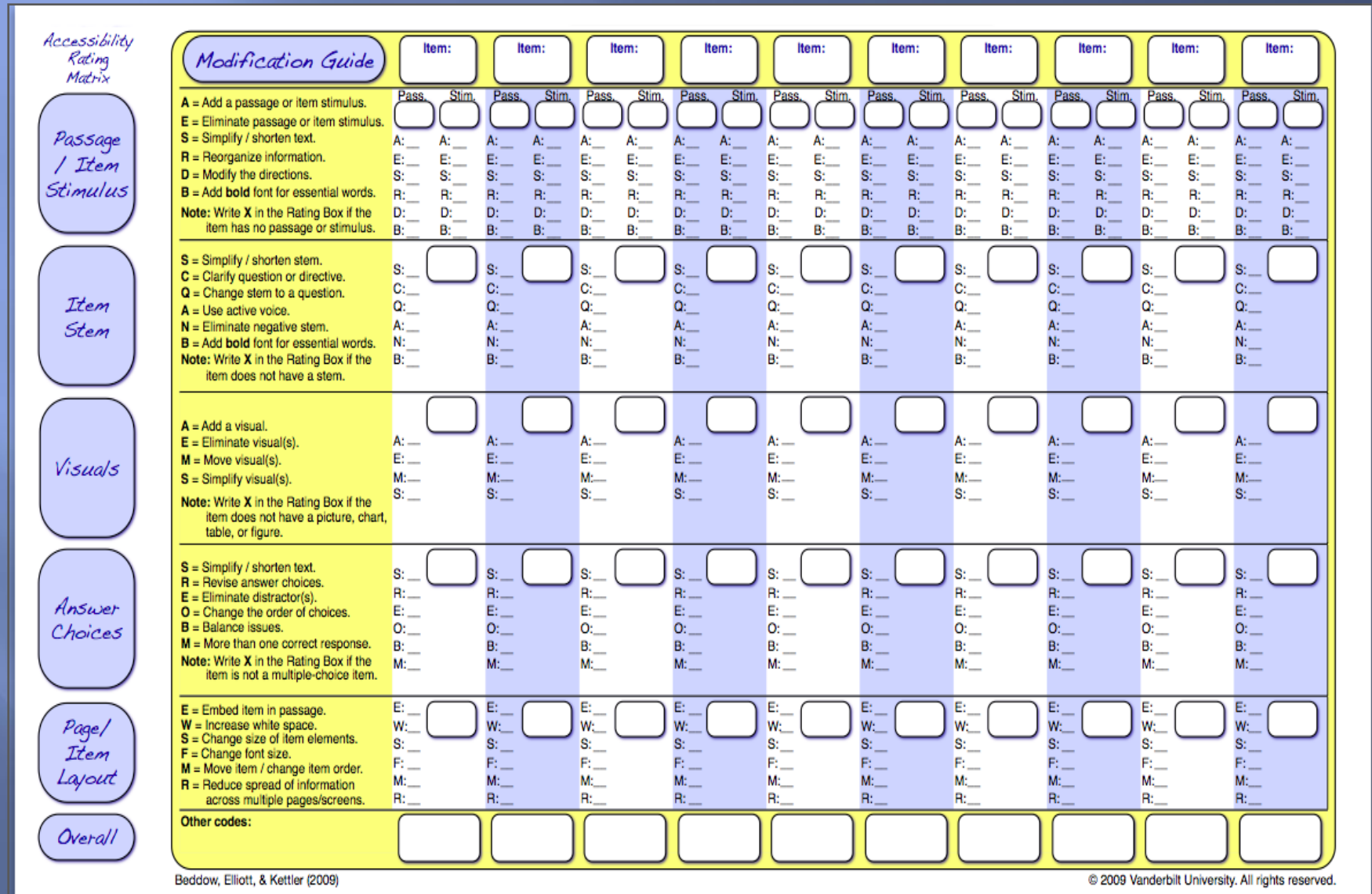
2
Maximally Accessible for Some Test-Takers

- Item contains content that is not essential for responding to the item, to the extent that it may be distracting or confusing to the test-taker.
- The wording of the item stem may cause some confusion as to what is required.
- Included visuals are unnecessary **and** potential distract the test-takers from essential item elements, **or** visuals are do not clearly depict the intended images or are unnecessarily complex.
- One or more distractors is implausible or absurd.
- Answer choices are unnecessarily complex or unbalanced with regard to length, content, and order.
- Rationale could be made for more than one correct response.
- Nonessential item elements in the page layout may draw test-taker attention away from essential content, or the test-taker must turn the page 2 or more times to respond to the item.

1
Inaccessible for Many Test-Takers

- The item contains a large amount of content that is not essential for responding to the item, to the extent that it is likely to confuse the test-taker.
- Stem is negatively worded, in passive voice, and/or it is not evident what is required.
- Included visuals are irrelevant and may cue test-taker to an incorrect response, **or** included visuals are likely to confuse the test-taker due to complexity or lack of clarity.
- Answer choices are unbalanced in a manner that may cue an incorrect response, contain more than one correct answer, and/or are implausible/absurd.
- Nonessential item elements in the page layout are likely to draw attention from essential information, or a large amount of essential information is presented across multiple pages/screens.

TAMI–ARM Record Form



Accessibility Rating Matrix

Modification Guide

Item: | Item: | Item: | Item: | Item: | Item: | Item: | Item: | Item: | Item:

Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim. | Pass. Stim.

Passage / Item Stimulus

A = Add a passage or item stimulus.
E = Eliminate passage or item stimulus.
S = Simplify / shorten text.
R = Reorganize information.
D = Modify the directions.
B = Add **bold** font for essential words.
Note: Write **X** in the Rating Box if the item has no passage or stimulus.

A: E: S: R: D: B: | A: E: S: R: D: B:

Item Stem

S = Simplify / shorten stem.
C = Clarify question or directive.
Q = Change stem to a question.
A = Use active voice.
N = Eliminate negative stem.
B = Add **bold** font for essential words.
Note: Write **X** in the Rating Box if the item does not have a stem.

S: C: Q: A: N: B: | ☐

Visuals

A = Add a visual.
E = Eliminate visual(s).
M = Move visual(s).
S = Simplify visual(s).
Note: Write **X** in the Rating Box if the item does not have a picture, chart, table, or figure.

A: E: M: S: | ☐

Answer Choices

S = Simplify / shorten text.
R = Revise answer choices.
E = Eliminate distractor(s).
O = Change the order of choices.
B = Balance issues.
M = More than one correct response.
Note: Write **X** in the Rating Box if the item is not a multiple-choice item.

S: R: E: O: B: M: | ☐

Page/ Item Layout

E = Embed item in passage.
W = Increase white space.
S = Change size of item elements.
F = Change font size.
M = Move item / change item order.
R = Reduce spread of information across multiple pages/screens.

E: W: S: F: M: R: | ☐

Overall

Other codes: ☐

Beddow, Elliott, & Kettler (2009)

© 2009 Vanderbilt University. All rights reserved.

Item Example – Original

32. Mr. Miyagi has instructed Daniel-San to spend the day washing all of his classic automobiles. If Daniel-San finishes the task, Mr. Miyagi will give him one of the cars for his birthday. Mr. Miyagi has 25 automobiles in all. If Daniel-San can wash 3 automobiles in 37 minutes, how many automobiles can he wash between noon and 4 o'clock in the afternoon?
A. 19 cars
B. 45 cars
C. 3 cars
D. 20 cars

33. Alexander and Eleanor have decided to spend the afternoon playing a game. They decide to play Foursquare. Alexander found a playground ball that was approximately 8 inches in diameter. Eleanor brought some playground chalk. Jimmy drew the court. The width and length of one square of the court are each 3 feet. What is the total area of the court in square feet?
A. 9 square feet
B. 12 square feet
C. 24 square feet
D. 36 square feet

34. Keanu is driving a bus. If the bus makes 5 stops, and it picks up an average of 7 people on each stop, about how many people will be on the bus when it reaches its destination?
A. 14 people
B. 49 people
C. 21 people
D. 35 people

35. Laverne organized a party to celebrate Shirley's 78th
All of their friends are invited. Arthur brought Quiche L
for dinner. If the radius of the quiche is 6 inches, what
circumference?
A. Approximately 38 inches
B. 12 inches
C. Approximately 113 inches
D. 60 inches

36. There are 145 invited guests to a luncheon organizec
Kiwanis club. If all of the attendees are over the age
what is the mean age of the people attending the lunc
A. 50
B. Over 50
C. Under 50
D. Approximately 60

37. Jorge wants to make a fruit salad for Thanksgiving. H
to the supermarket and bought 4 grapefruit, 6 kiwi, 12
4 pears, 11 bananas, and 3 oranges. Bananas and kiv
up what percentage of the total number of fruit Jorge b
A. 12.5 percent
B. 25 percent
C. 42.5 percent
D. Over 50 percent

38. September, April, June, and November all consist of 3
February consists of either 28 or 29 days, depending
whether it is a leap year. How many months have mor
30 days?

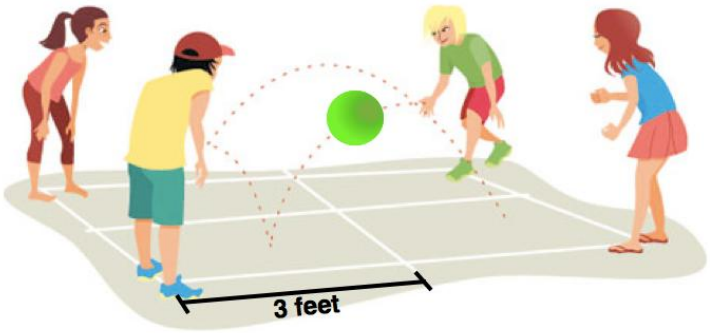
Domain: Math
Words: 75
Readability: 5.5

TAMI ARM rating:
Inaccessible for many test-takers (Rating: 1)

Target construct:
Use knowledge of 2-dimensional shapes to solve real-world calculation problems (Gr. 5)

Item Example – Modified

33 Foursquare is a game that is played with a ball on a court that is made up of **4 congruent squares**. Look at the picture below.



The kids are playing Foursquare. If one square is 3 feet long and 3 feet wide, what is the **total area** of the Foursquare court?

- ◯ 9 square feet
- ◯ 24 square feet
- ◯ 36 square feet

Domain: Math
Words: 60 (-20%)
Readability: 3.2 (-2.3)

TAMI ARM rating:
Maximally Accessible
for nearly all test-takers
(Rating: 4)

Target construct:
Use knowledge of 2-dimensional shapes to solve real-world calculation problems (Grade 5)

Modifications Benefited all Groups

Cohen's d for Reading and Mathematics by Group and Modification [1]

	Modified	Modified with Reading Support	Reading Support Over Modified
Reading			
SWOD	.37	.38	.01
SWD-NE	.38	.49	.11
SWD-E	.40	.50	.11
Total	.38	.46	.07
Mathematics			
SWOD	.15	.20	.05
SWD-NE	.21	.25	.05
SWD-E	.26	.31	.04
Total	.21	.25	.05

[1] To determine the effect sizes for the Modified and Modified with Reading Support condition, the Original condition served as the point of comparison.

Figure 3. Mean test scores by group and condition in reading.

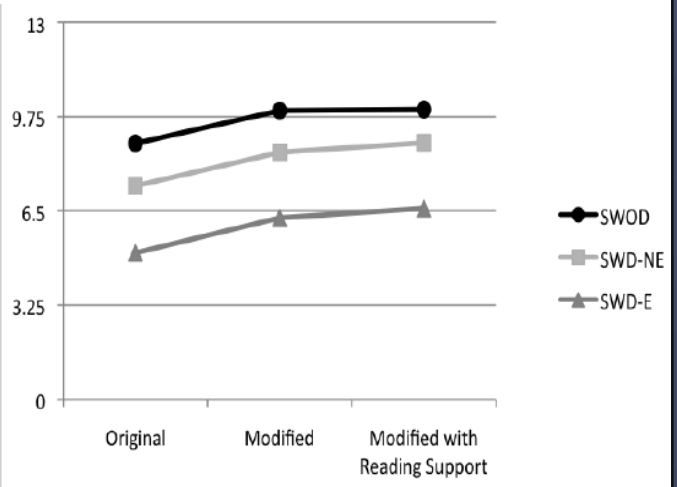
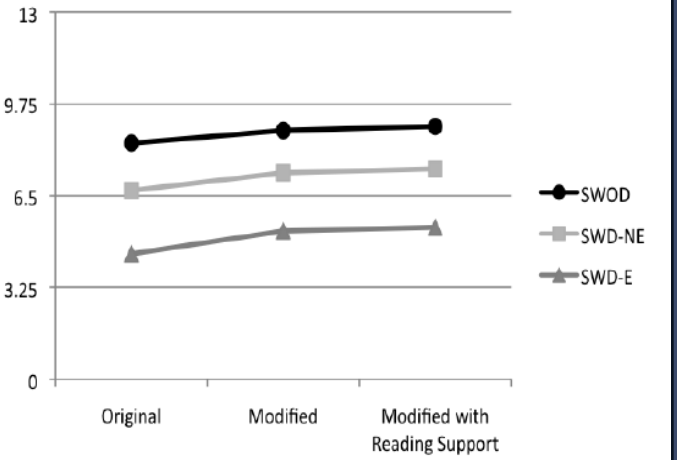


Figure 4. Mean test scores by group and condition in mathematics.



Elliott, et al. (2010), *Exceptional Children*

Evidence Indices for Desired Changes

If Desired Outcomes	Indices of Change
Increase accessibility	TAMI Overall ARM rating
Decrease item difficulty	p value or percent correct
Increase item discrimination	D or point-biserial correlation
Increase reliability estimates	KR-20 or coefficient alpha
Reduce readability level	Flesch-Kincard grade level
Maintain alignment w/content stds.	Judged to be aligned when using approved alignment method
Maintain DOK for all items	DOK level indicator
Reduce need for accommodations	Number of accommodations
Improve students' perceptions of tests & motivation to engage in testing	Cognitive ease ratings , self-reports, time

It would facilitate communication & advance our science if we all reported at least a common set of item/test indices!

Lessons Learned from AA-MAS Item Modification Research

1. One characteristic shared by many in the AA-MAS eligible group is slow reading. Thus, reducing number of words to be read facilitates persistence, engagement, and comprehension.
2. Modified items that are less complex, but still aligned with grade level content standards and measuring the same DOK, can result in improved performance for all students.
3. Easy and effective modifications include highlighting the question asked, simplifying language, and reducing the number of answer choices. These elements often shorten a given item, thus also allowing for more items within the same general time frame.
4. Modified items can result in equal or better measurement precision for the eligible population.

Conclusion from Item Modification Studies: *Less is More*…

…or at least, *Less is* ***equal*** *(in terms of measurement precision) and* ***simpler*** *(in terms of reading and cognitive load) and allows eligible students a better opportunity to experience success, so less is also* ***better*** for all students.

Tests are Designed to Measure What Students Know and the Effects of their Instruction, but ……..

What if instruction on the content covered by the test item has not been provided?

What if some students received instruction and others have not received instruction on the content tested?

- A number of students and also their teachers told us that they had NOT been taught content measured on their state test.
- Test developers have usually tried to address these concerns by conducting alignment studies, using tools/methods like the SEC or Webb's Alignment method. Note these methods are insensitive to individual students.

Alignment is a poor proxy measure for Opportunity to Learn!

Initial Method for Documenting OTL

Curricular Experiences Survey/4th grade
Cognitive Lab Study
CMAADI-AZ Project

Student's Name:

Directions: For the student listed above, indicate the extent of content coverage (# of lessons provided during this current school year) for each of the following objectives.

NC = *No coverage* (Not an objective addressed during this school year)
P = *Plan to cover* this objective later in this school year
1 = *Minimal coverage* (1-5 lessons to this point in the school year)
2 = *Moderate coverage* (6 -10 lessons to this point in the school year)
3 = *Extensive coverage* (11 or more lessons to this point in the school year)
4 = *Intensive, systematic coverage* (daily/nearly daily instruction throughout the school year)

Reading/Language Arts

Objective	No Coverage	Plan to Cover	Coverage Provided			
1. Using context to determine the relevant meaning of a word.	NC	P	1	2	3	4
2. Indentifying the main idea and supporting details.	NC	P	1	2	3	4
3. Describing the historical and cultural aspects found in cross-culture works.	NC	P	1	2	3	4
4. Determining the author's main purpose.	NC	P	1	2	3	4
5. Determining the author's position regarding a particular idea, subject, concept, or object.	NC	P	1	2	3	4

CMAADI AZ Cog Lab Study, 2008

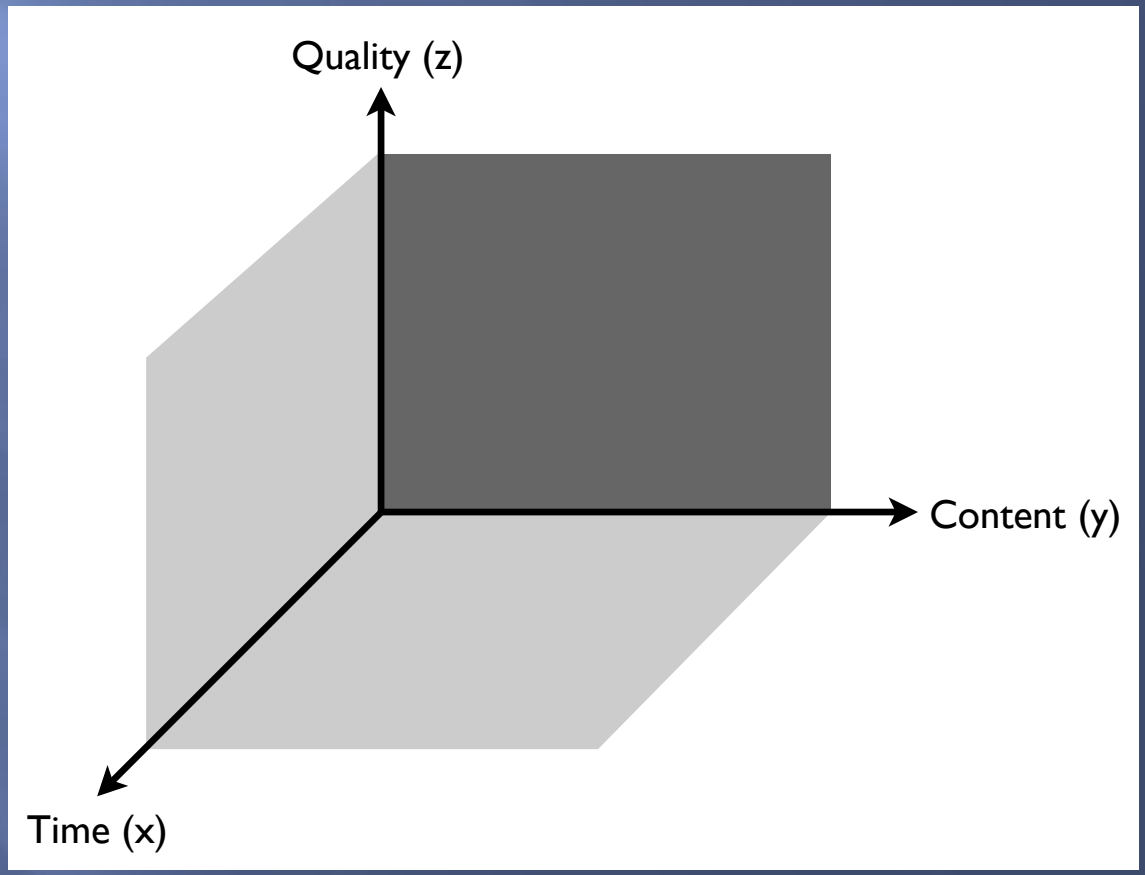
Opportunity to Learn (OTL)

Opportunity to learn is defined as: *the degree to which a teacher dedicates instructional time and content coverage to the intended curriculum objectives emphasizing high-order cognitive processes, evidence-based practices, and alternative grouping formats.*

This definition is the conceptual foundation for the indices measured by the Instructional Learning Opportunity Guidance System (**MyiLOGS; Kurz, Elliott, & Shrago, 2009**), an online measure developed in a recently completed **USDE Enhanced Assessment Grant (Award # S368A090006).**

Time, Content, & Quality all Matter!

Unified Model of OTL



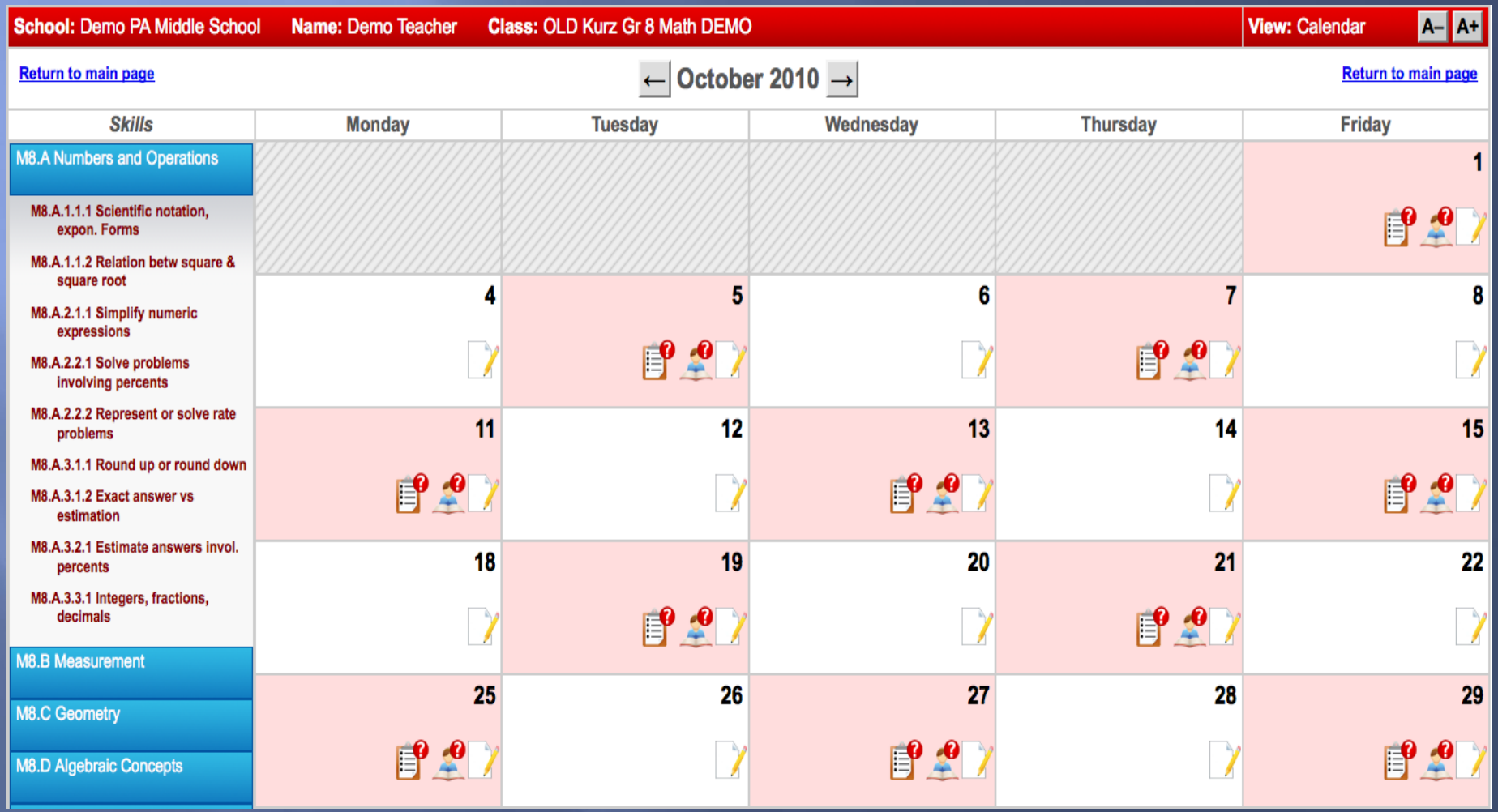
MyiLOGS: A Measure of OTL

• **My instructional Learning Opportunities Guidance System** (MyiLOGS; Kurz, Elliott, & Shrago, 2009) allows teachers to document their planned and enacted instruction along their state-specific intended curriculum. Created as part of an USDE Enhanced Assessment Grant to Pennsylvania Dept. of Education.

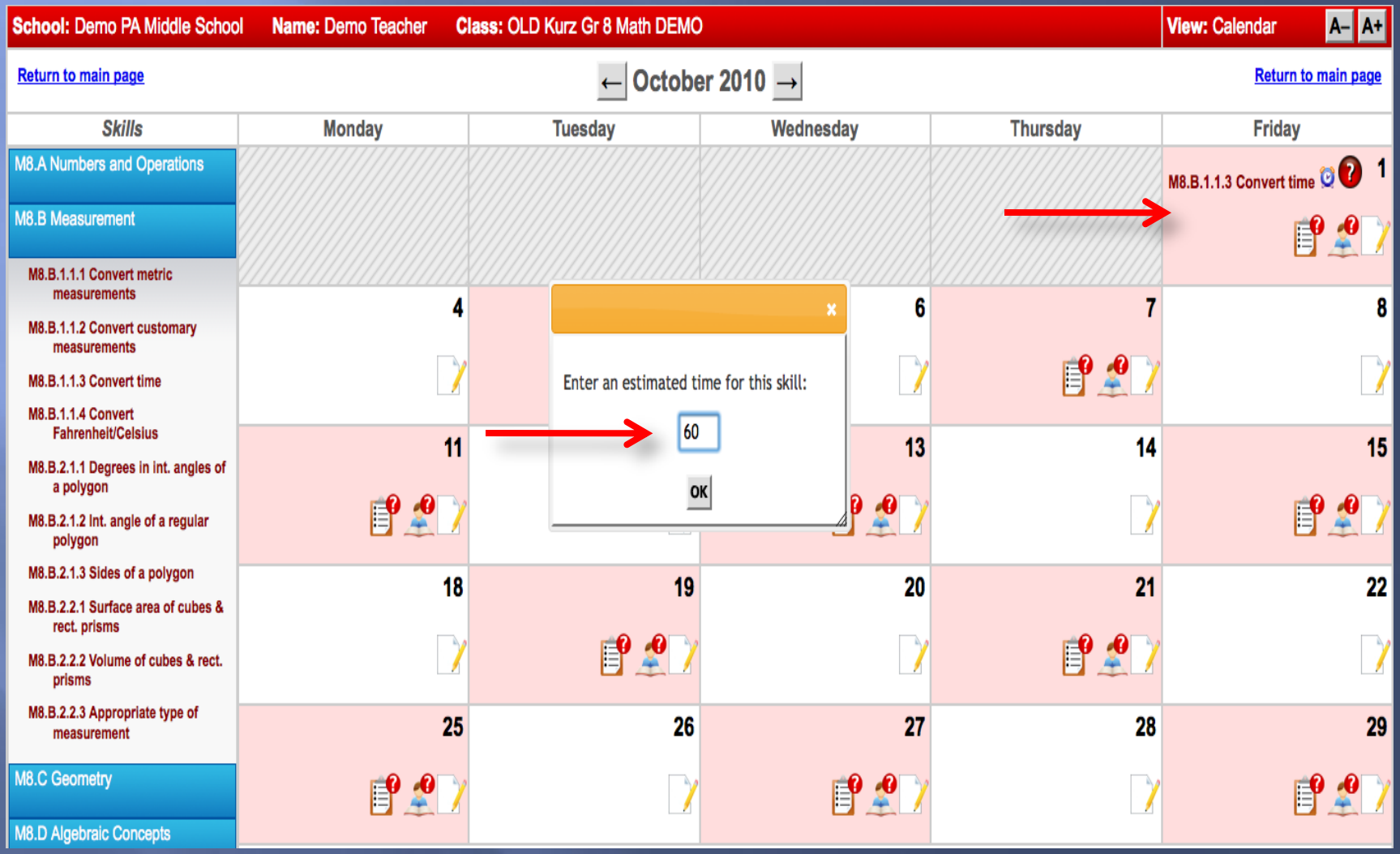
• OTL is documented along **three key dimensions** (i.e., time, content, and quality) at the **classroom and individual student level**. MyiLOGS captures:

- Coverage of state-specific subskills
- Time spent on each subskill
- Cognitive expectations for student learning
- Use of evidence-based instructional practices
- Use of instructional grouping formats
- Student engagement

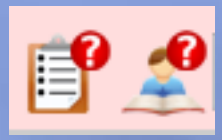
Advancing the Measurement of OTL:
The MyiLOGS Core Curriculum Calendar



The MyiLOGS Calendar Example



MyiLOGS: Enacted Details



School: Demo PA Middle School
Teacher: Demo Teacher
Class: TRAINING Kurz Gr 8 Math

Date: Wed., Sep 15

Class Enacted | Student Enacted

Return to Calendar and add / delete skills | Save time allocation | Clear values

Estimated Time Allocation Across Cognitive Process Dimensions for: TRAINING Kurz Gr 8 Math

Skill	Remember	Understand/Apply	Analyze/Evaluate	Create	Sum	Calendar Minutes
M8.A.3.1.1 Round up or round down	0	20	20	20	60	60
Time not available for instruction						
				Update Totals	Total: 60	60

Estimated Time Allocation Across Instructional Practices for: TRAINING Kurz Gr 8 Math

Teacher Actions	Individual	Small Group	Whole Class	Sum
Provided explicit instruction	0	0	20	20
Provided visual representations	0	0	20	20
Taught problem solving strategies	0	0	20	20
Elicited 'think aloud'	0	0	0	0
Provided guided feedback	0	0	0	0
Used independent practice	0	0	0	0
Other instructional practices	0	0	0	0
Conducted assessment	0	0	0	0
Time not available for instruction				0
		Update Totals	Calendar Total: 60	60

MyiLOGS Observation Record

Teacher ID:	Date:	Class:	Time:

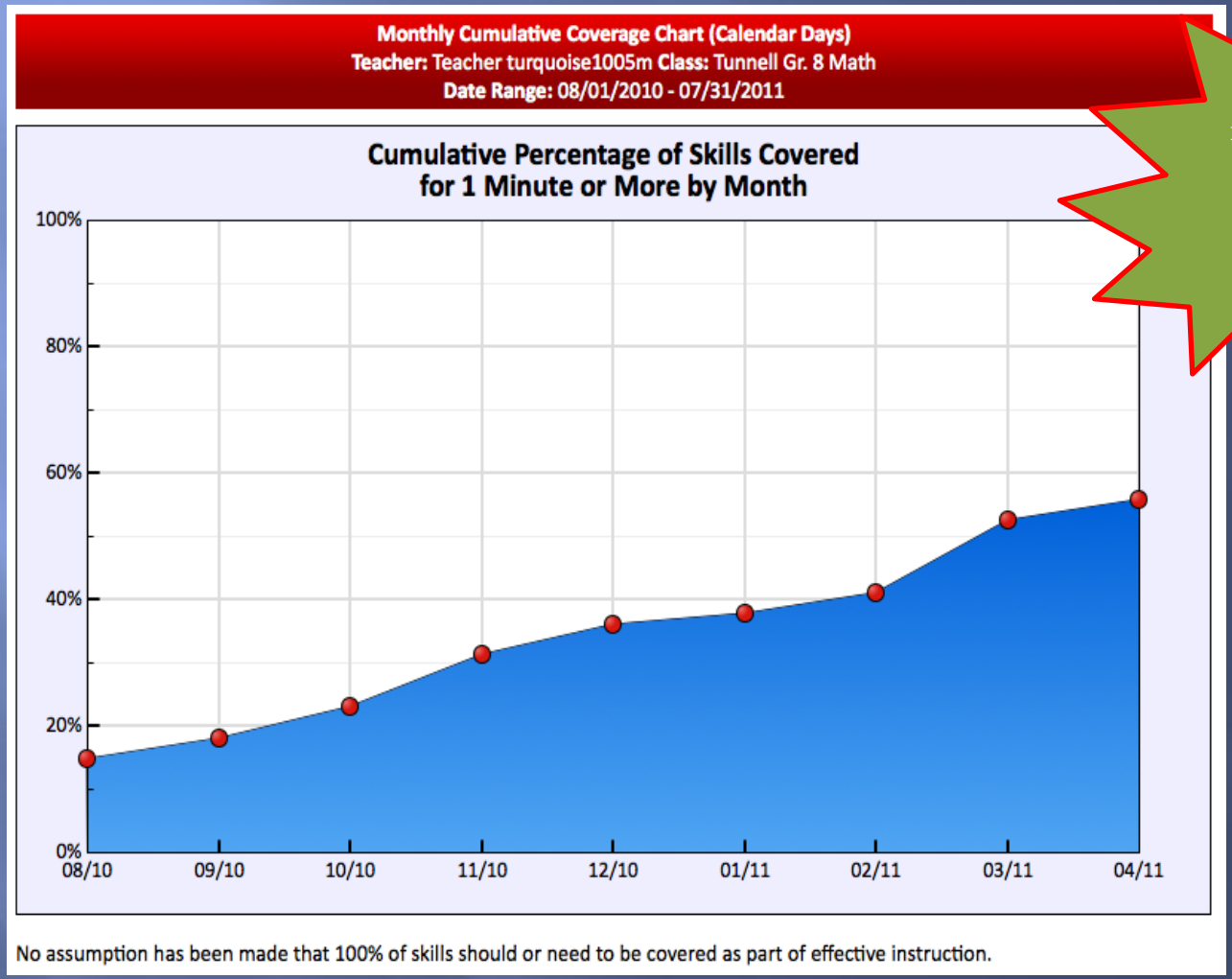
❶ Record in 1-min intervals. ❷ Use tally marks (𝍸) to record the **student expectation** and **teacher action** that occupied the **majority of time** during the 1-min interval.

Skills	Attend Listen, focus, pay attention	Remember Recognize, identify, recall	Understand/Apply Interpret, exemplify, classify, summarize, infer, compare, explain / Execute, implement, use	Analyze/Evaluate Differentiate, organize, integrate, attribute / Check, test, critique, judge	Create Generate, hypothesize, plan, design, produce	Comments
Time not available for instruction						

Teacher Actions	Individual Action is focused on single individuals	Small Group Action is focused on small groups	Whole Class Action is focused on entire class	Comments
Provided Direct Instruction Teacher presents issue, discusses or models a solution approach, and engages students with approach in similar context.				
Provided Visual Representations Teacher uses visual representations to organize information, communicate attributes, and explain relationships.				
Asked Questions Teacher asks questions to engage students and focus attention on important information.				
Elicited Think Aloud Teacher prompts students to think aloud about their approach to solving a problem.				
Used Independent Practice Teacher allows students to work independently to develop and refine knowledge and skills.				
Provided Guided Feedback Teacher provides feedback to students on work quality, missing elements, and observed strengths.				
Provided Reinforcement Teacher provides reinforcement contingent on previously established expectations for effort and/or work performance.				
Assessed Student Knowledge Teacher uses quizzes, tests, student products, or other forms of assessment to determine student knowledge.				
Other Instructional Practices Any instructional practices not captured by the aforementioned key instructional practices. You can use the class notes to leave additional details.				
Time not available for instruction				

Summative Class Engagement	○ 0% ○ <50% ○ 50%-80% ○ >80%	Summative Goal Attainment	○ 0% ○ <50% ○ 50%-80% ○ >80%

MyiLOGS: Instructional Reports

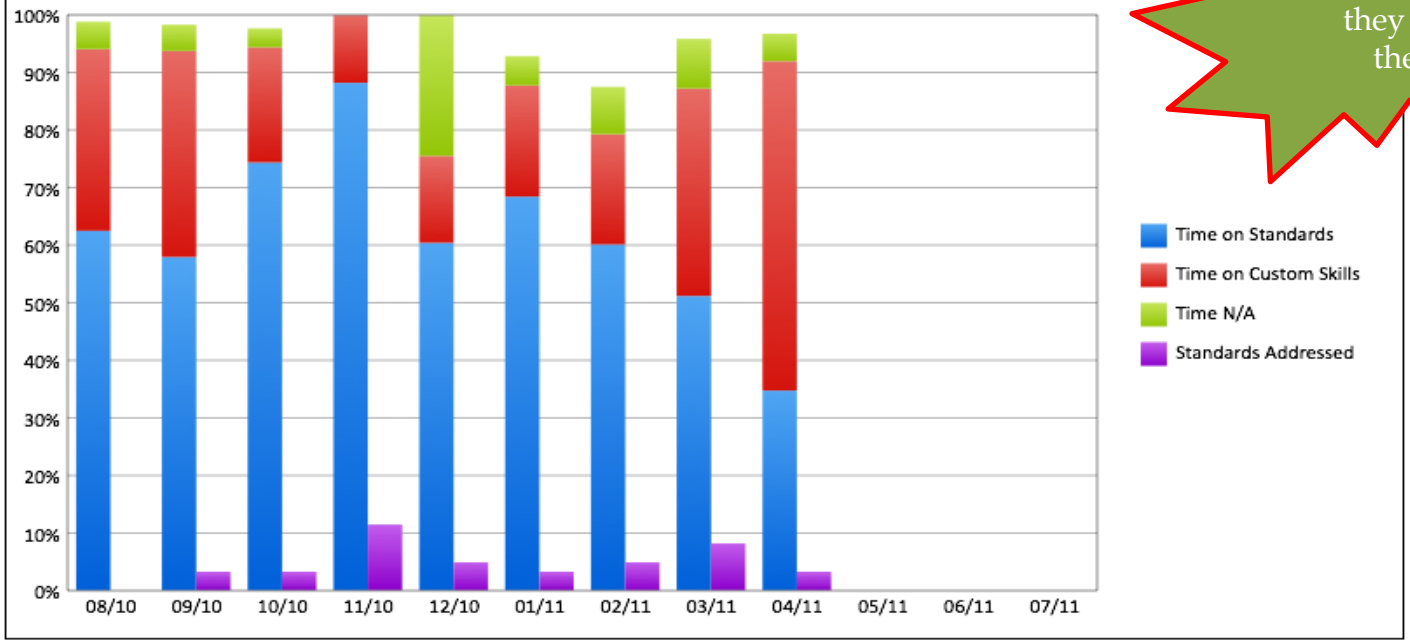


Teachers get individualized reports when they want them!

MyiLOGS: Instructional Reports

Teachers get individualized reports when they want them!

OTL Summary Indices: Monthly Within-Subject Comparison Chart (Calendar and Sample Days)
Teacher: Teacher turquoise1005m **Class:** Tunnell Gr. 8 Math
Date Range: 08/01/2010 - 07/31/2011



	08/10	09/10	10/10	11/10	12/10	01/11	02/11	03/11	04/11	05/11	06/11	07/11
Time on Standards*	63%	58%	74%	88%	60%	68%	60%	51%	35%	—	—	—
Time on Custom Skills*	32%	36%	20%	12%	15%	19%	19%	36%	57%	—	—	—
Time on N/A*	5%	5%	3%	0%	24%	5%	8%	9%	5%	—	—	—
Standards Addressed	0%	3%	3%	11%	5%	3%	5%	8%	3%	—	—	—

*All three time indices should sum to 100% of the scheduled class length. Please review discrepancies, as they may reflect reporting errors or legitimate adjustments in school time that may occur occasionally.

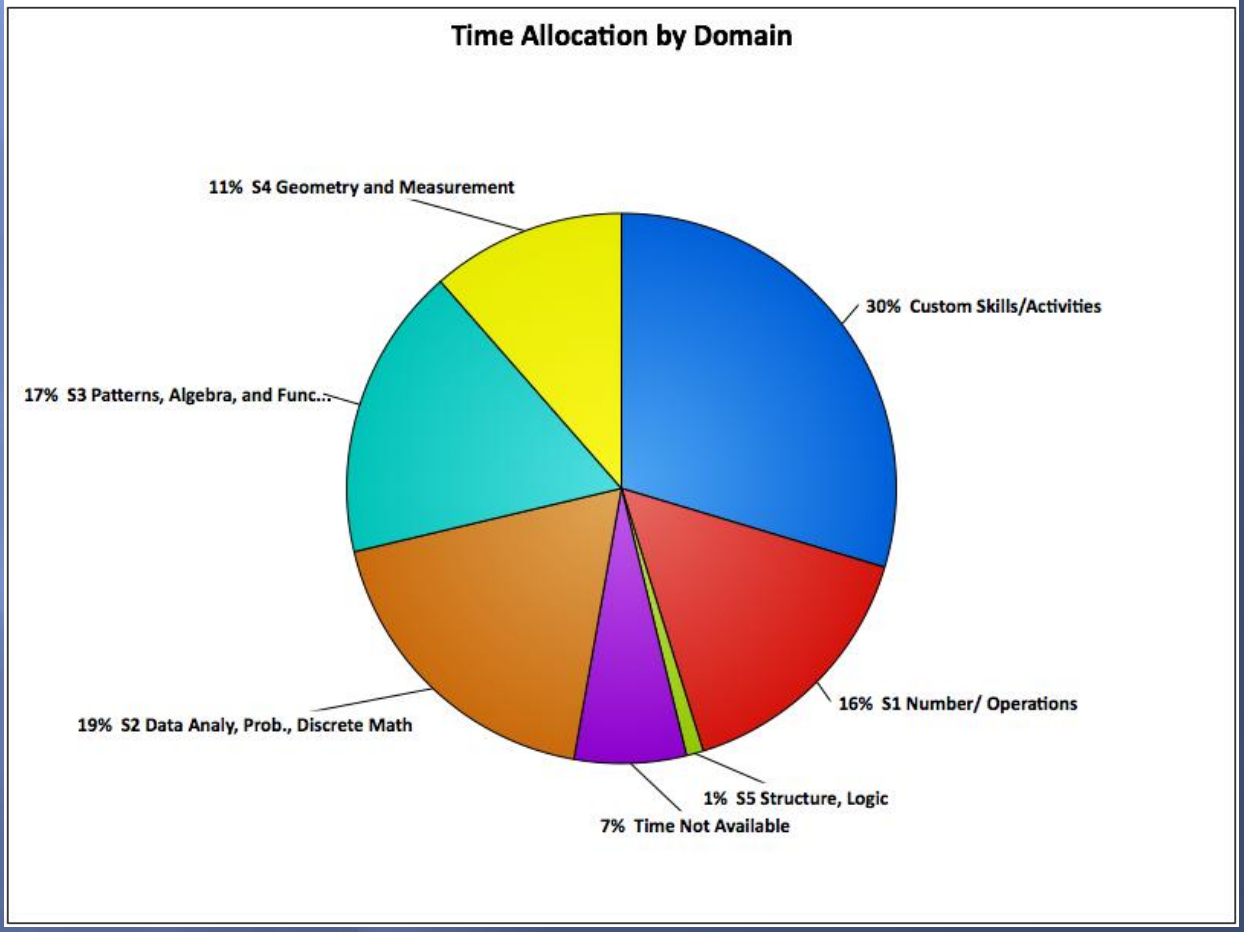
MyiLOGS: Instructional Reports

Detailed Content Coverage Bar Chart (Calendar Days)
Teacher: Teacher turquoise1005m **Class:** Tunnell Gr. 8 Math
Date Range: 08/01/2010 - 07/31/2011

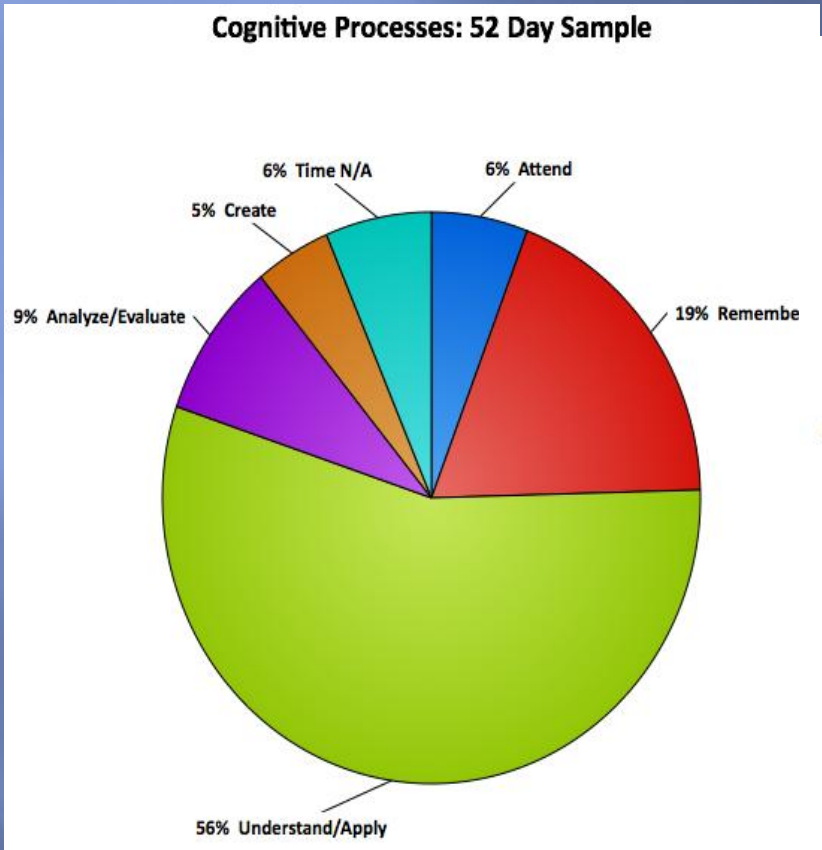
Skills	% of Total Time	
S1 Number/ Operations	**15.6%**	**33 hrs 20 mins**
S1C1PO1 Compare/order	0.6%	1 hrs 15 mins
S1C1PO2 Classify rational/irrational	1.5%	3 hrs 15 mins
S1C1PO3 model read numbers	1.5%	3 hrs 15 mins
S1C1PO4 model/solve absolute value	0.5%	1 hrs 10 mins
S1C2PO1 Factors/multiples/prime	1.6%	3 hrs 30 mins
S1C2PO2 Rational number effects	1.0%	2 hrs 10 mins
S1C2PO3 Percent inc., dec, simple interest	3.1%	6 hrs 45 mins
S1C2PO4 Std/scientific notation conver.	1.4%	3 hrs 5 mins
S1C2PO5 Simplify expression	3.4%	7 hrs 20 mins
S1C3PO1 Estimate1		
S1C3PO2 Estimate on number line	0.7%	1 hrs 35 mins
S2 Data Analy, Prob., Discrete Math	**18.5%**	**39 hrs 42 mins**
S2C1PO1 Use displays, box-whisker, scatterplot	5.2%	11 hrs 12 mins
S2C1PO2 Inferences, 2 data sets	1.2%	2 hrs 40 mins
S2C1PO3 Summary-shape of distribution		
S2C1PO4 Bias, effective presentation	0.2%	0 hrs 30 mins
S2C1PO5 Evaluate design		
S2C2PO1 Theoretical/experimental	3.2%	6 hrs 50 mins
S2C2PO2 Compare outcome/prediction	2.3%	5 hrs 0 mins
S2C2PO3 Sample space for dep/indep	1.5%	3 hrs 15 mins
S2C3PO1 Counting-order,repetition		
S2C3PO2 Counting-factorial notation	2.4%	5 hrs 5 mins
S2C4PO1 Solve graph problems	2.4%	5 hrs 10 mins
S3 Patterns, Algebra, and Functions	**17.3%**	**36 hrs 59 mins**

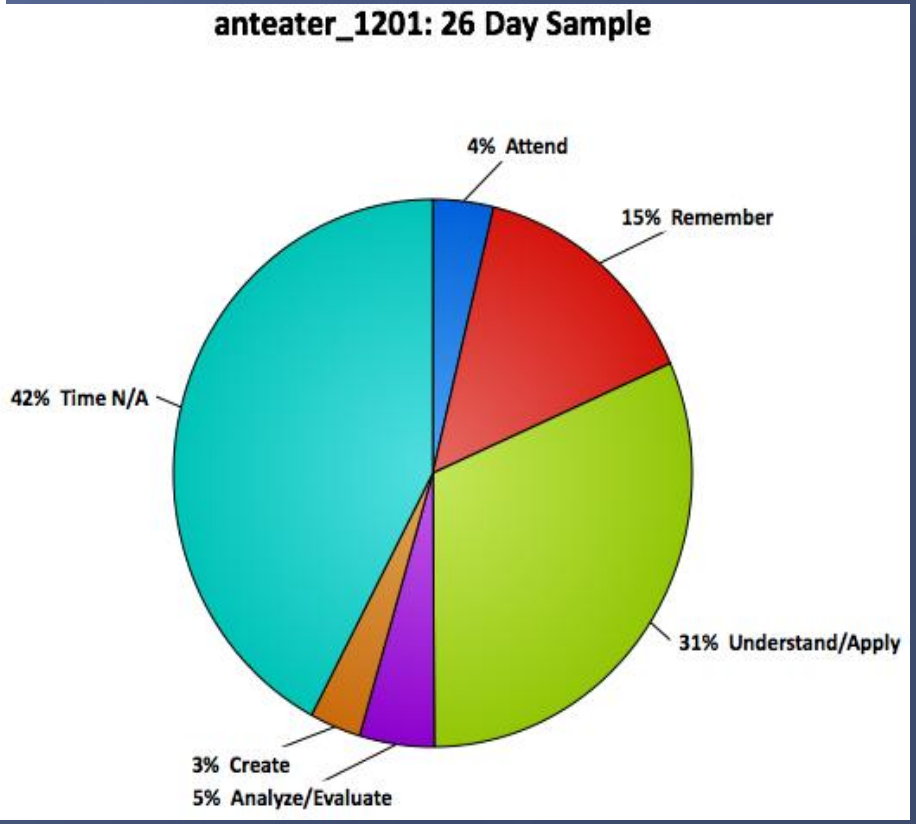
MyiLOGS: Instructional Reports

Broad Content Domain Coverage Pie Chart (Calendar Days)
Teacher: Teacher turquoise1005m **Class:** Tunnell Gr. 8 Math
Date Range: 08/01/2010 - 07/31/2011



Comparison of Class and Student: Cognitive Processes





Key Findings Regarding OTL

- Initial OTL data from 3 states indicated that general and special education teachers spent about 66% of their allocated class time on teaching the academic standards of the general curriculum, another 25% on custom skills/activities, and about 5% on non-instructional activities/tasks. Teachers covered approximately 68% of the academic standards based on an average of about 151 school days before their state test.
- Teachers placed greater emphasis on higher-order thinking skills in general education classrooms than in special education classrooms.
- Differences between general and special education teachers related to Time on Standards and Content Coverage indicated effect sizes above .50. Given that students in both types of classes were held to the same general curriculum standards irrespective of educational setting, it is problematic that teachers in special education classes provided less instructional time and coverage of the academic standards.

Major Conclusion from OTL Studies

"Based on this sample's general education classrooms, which represented a full inclusion model, students with disabilities experienced less time on standards, more non-instructional time, and less content coverage compared to their class. … At least for students with disabilities nested in general education classrooms, OTL appears to be a differentiated opportunity structure. …the instructional differences do not indicate equal or equitable OTL for students with disabilities. Given their disability-related characteristics, students with disabilities may need at least as much OTL, if not more, than their peers without disabilities. However, the current findings suggest the exact opposite; if replicable, these data would pose serious instructional challenges for teachers and hold profound implications for policy makers focusing on academic proficiency and growth without consideration for the instructional inputs and processes that affect student outcomes."

(Kurz, Elliott, Lemons, Zigmond, Kloo, & Kettler, 2012)

Implications of OTL Research for Test Developers

- Studies must be conducted to establish the reasonableness of validity claims about relationships between instruction and test scores.
- Teachers need substantial support to meaningfully cover the intended general curriculum with all students, in particular those with disabilities. Many students with disabilities will need 30 to 40 more days of class time annually to have equitable OTL.
- Alignment is important, but current alignment measures are insensitive to individual students instruction. Thus, alignment is a poor indicator of OTL for students with disabilities.

1. How can tests be modified to improve the measurement of knowledge and skills of students with persistent academic difficulties and who would qualify to take an AA-MAS, if available?

2. How can testing conditions be improved to increase the likelihood that all students have had the opportunity to learn the content standards that the tests are designed to measure?

Our research has provided some practical tools that can help answer these questions and offer students with disabilities more opportunities to demonstrated what they have learned.

Make research matter! Use this knowledge to advance assessment and instructional practices for all students.

Key References

- Clark, R., Nguyen, F., & Sweller, J. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. San Francisco: Pfeiffer.
- Elliott, S. N., Kettler, R. J., Beddow, P. A., & Kurz, A. (Eds.). (2011). Handbook of accessible achievement tests for all students: Bridging the gaps in policy, research, and practice. New York: Springer
- Elliott, S.N., Kettler, R.J., Beddow, P.A., Kurz, A., Compton, E., McGrath, D., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M., Bolt, D., & Roach, A.T. (2010). Effects of using modified items to test students with persistent academic difficulties. Exceptional Children, 76 (4), 475-495.
- Kettler, R.J., Elliott, S.N., & Beddow, P.A. (2009). Modifying achievement test items: A theory-guide and data-based approach for better measurement of what students with disabilities know. Peabody Journal of Education, 84, 529-551. DOI: 10.1080/01619560903240996.
- Kettler, R.J., Rodriguez, M.R., Bolt, D.M., Elliott, S.N., Beddow, P.A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. Applied Measurement in Education, 24, 1-25.
- Roach, A.T., Beddow, P., Kurz, A., Kettler, R.J., & Elliott, S.N. (2010). Incorporating student input in developing alternate assessments based on modified achievement standards. Exceptional Children, 77 (1), 61-84.
- Rodriguez, M.C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educational Measurement: Issues and Practice, *24*(2), 3-13.

ASU LEARNING SCIENCES INSTITUTE
ARIZONA STATE UNIVERSITY

Thank you very much for your time.

Your comments and feedback are welcome.

Steve_elliott@asu.edu

Resources mentioned can be found at:

- http://peabody.vanderbilt.edu/LSI_Projects/CAAVES_Project_Home.xml
- http://peabody.vanderbilt.edu/LSI_Projects/C-MAADI_Project_Home.xml
- http://peabody.vanderbilt.edu/TAMI.xml
- www.myilogs.com

Innovative Item Access Features for Students with Disabilities in Virginia

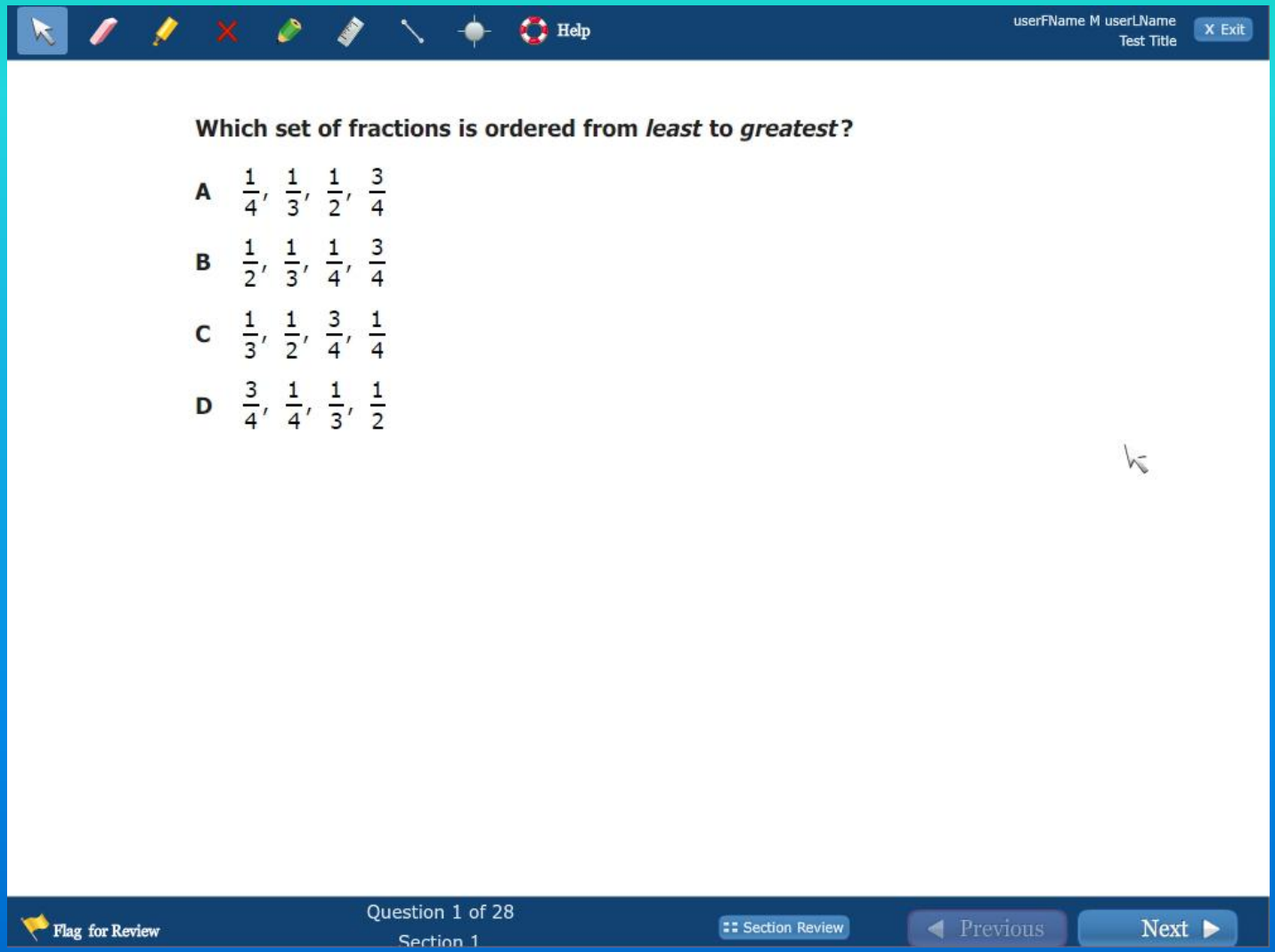
Shelley Loving-Ryder
Assistant Superintendent for Student Assessment and School Improvement

May 22, 2012

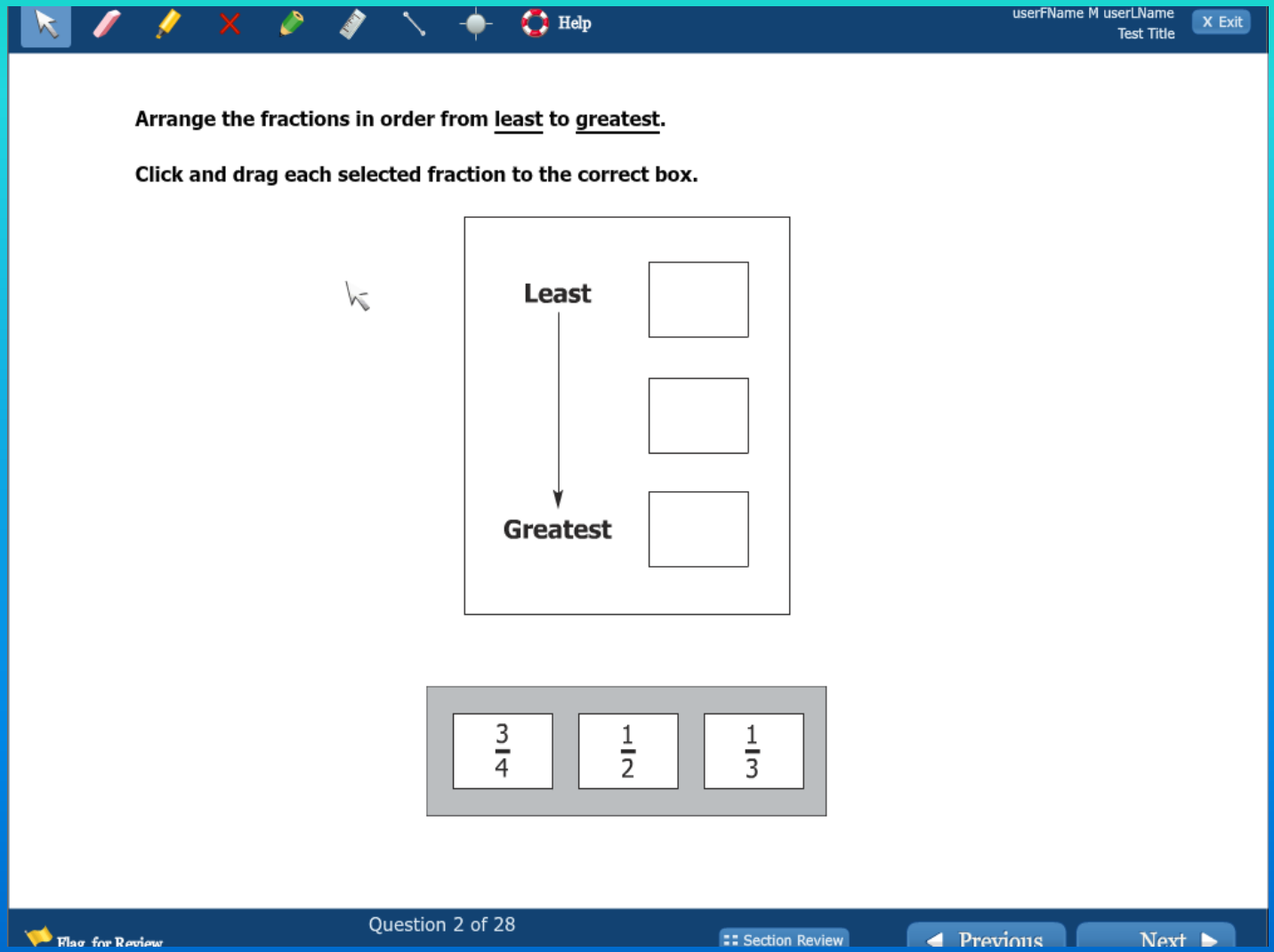
Virginia Modified Achievement Standards Test (VMAST)

- Grant to develop modified achievement standards assessment for grade 8 reading and mathematics
- Added supports and simplifications recommended by educators to existing online test items
- Expanded VMAST to include grades 3-8 mathematics and Algebra I (operational in spring 2012) and reading for grades 3-8 and end-of-course (operational in spring 2013)

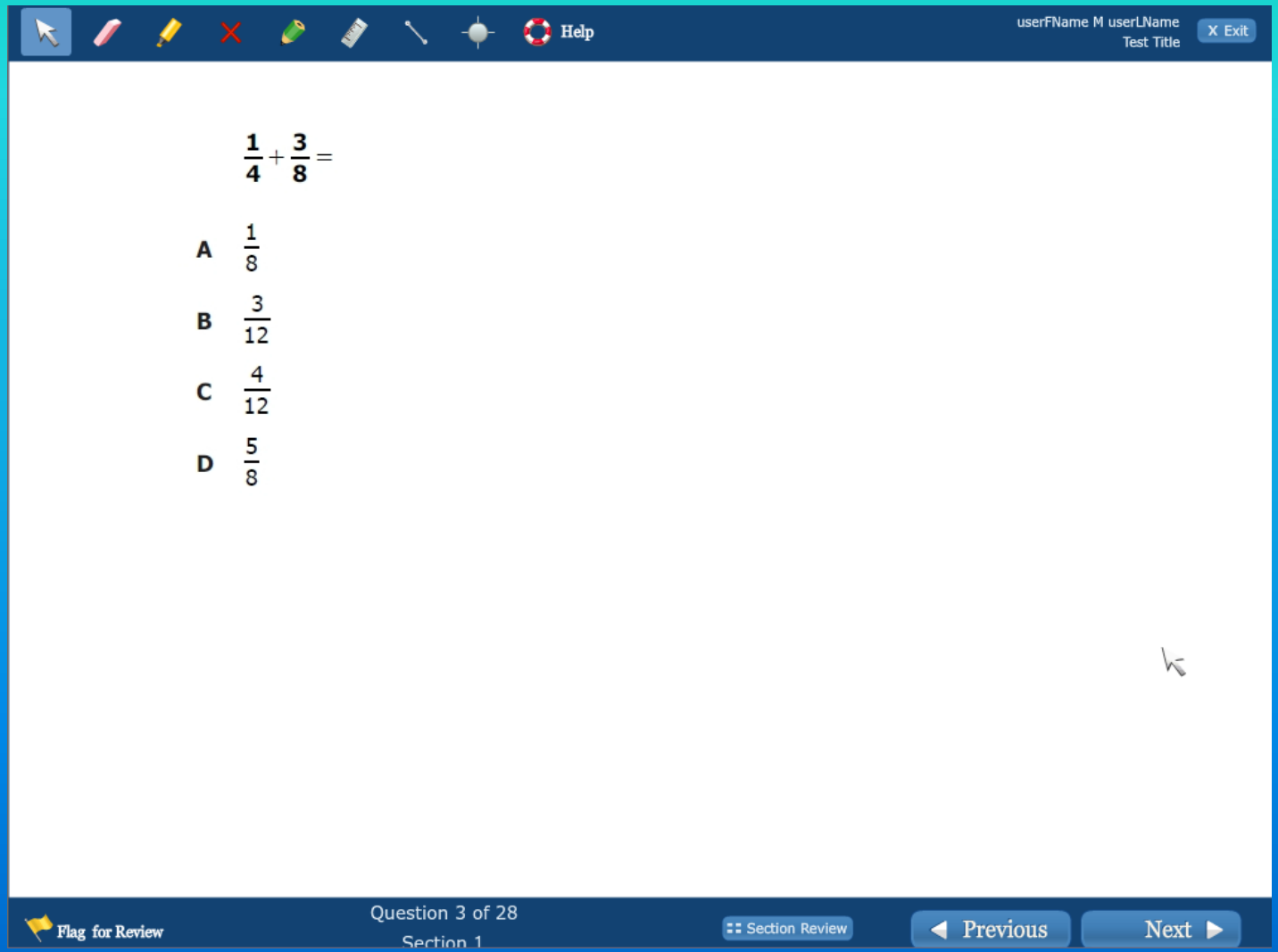
Grade 4 Mathematics Standards of Learning (SOL) Item



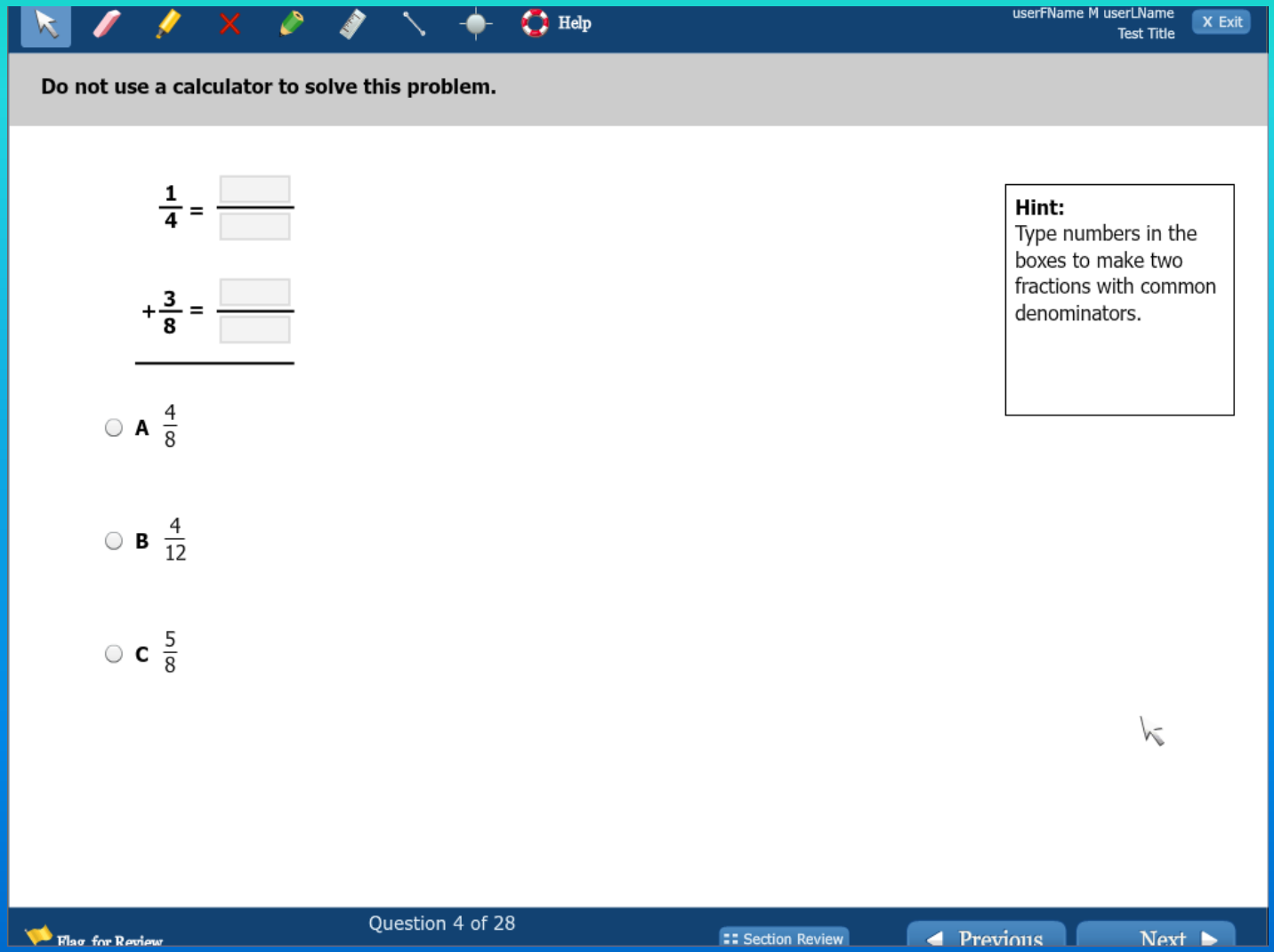
Grade 4 Mathematics
Virginia Modified Achievement Standards Test (VMAST) Item



Grade 4 Mathematics SOL Item



Grade 4 Mathematics
VMAST Item



Grade 5 Mathematics SOL Item

Help

userFName M userLName
Test Title

X Exit

When section 3 is removed from the parallelogram shown, which of the following *best* describes the new figure?

A Kite
B Rhombus
C Trapezoid
D Rectangle

Flag for Review

Question 5 of 28

Section Review

Previous

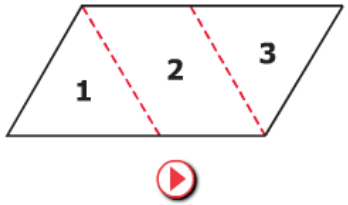
Next

Grade 5 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit



Hint:
Click on ▶ to see the individual shapes.

The figure shown will be cut along the dashed line segments. What shapes will result?

- **A** Two triangles and a rectangle
- **B** Two triangles and a parallelogram
- **C** A parallelogram, a triangle, and a kite

Flag for Review

Question 6 of 28

Section Review

◀ Previous

Next ▶

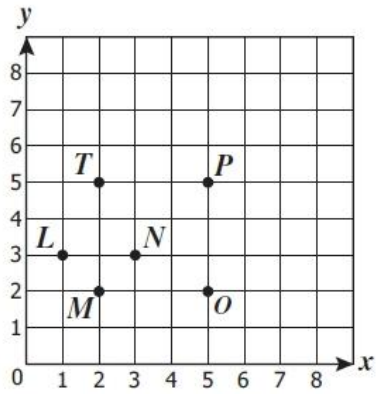
Grade 6 Mathematics SOL Item

Help

userFName M userLName
Test Title

X Exit

The picture shows six points on a grid.



Which three points can be connected to form a right angle?

A Points T, L, and N

B Points L, P, and T

C Points N, O, and P

D Points M, O, and P

Flag for Review

Question 7 of 28

Section Review

Previous

Next

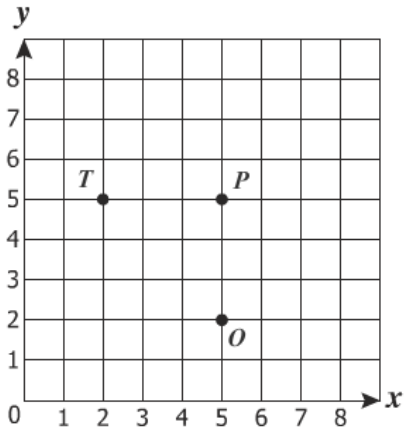
Grade 6 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit

Which point best represents the ordered pair (5 , 2)?



Hint:
Place the pointer tool over each answer choice to highlight the point and lines of intersection. Then select your answer.

- **A** Point O
- **B** Point P
- **C** Point T

Flag for Review

Question 8 of 28

Section Review

Previous

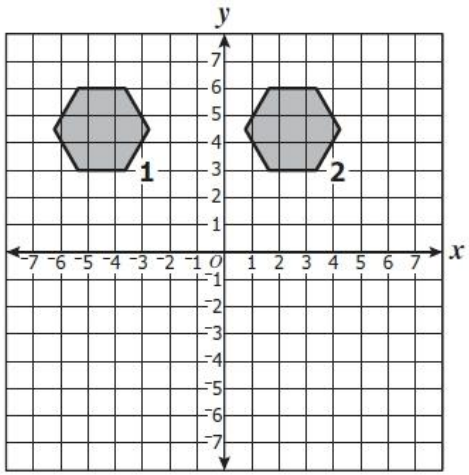
Next

Grade 7 Mathematics SOL Item

Help

userFName M userLName
Test Title

X Exit



Which is most likely the type of transformation that takes place from Figure 1 to Figure 2 on the coordinate grid above?

A Rotation about the origin
B Dilation
C Translation
D Reflection across the y-axis

Flag for Review

Question 9 of 28
Section 1

Section Review

◀ Previous

Next ▶

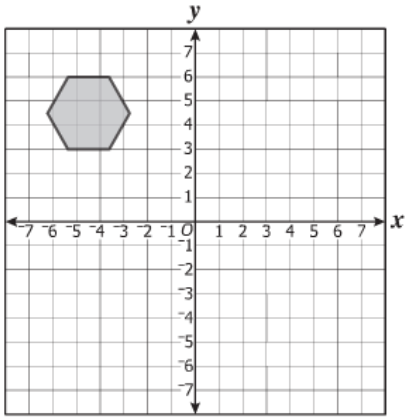
Grade 7 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit

Click on ▶ to show the transformation.



What type of transformation was shown?

- **A** Rotation about the origin
- **B** Translation
- **C** Reflection across the y-axis

Flag for Review

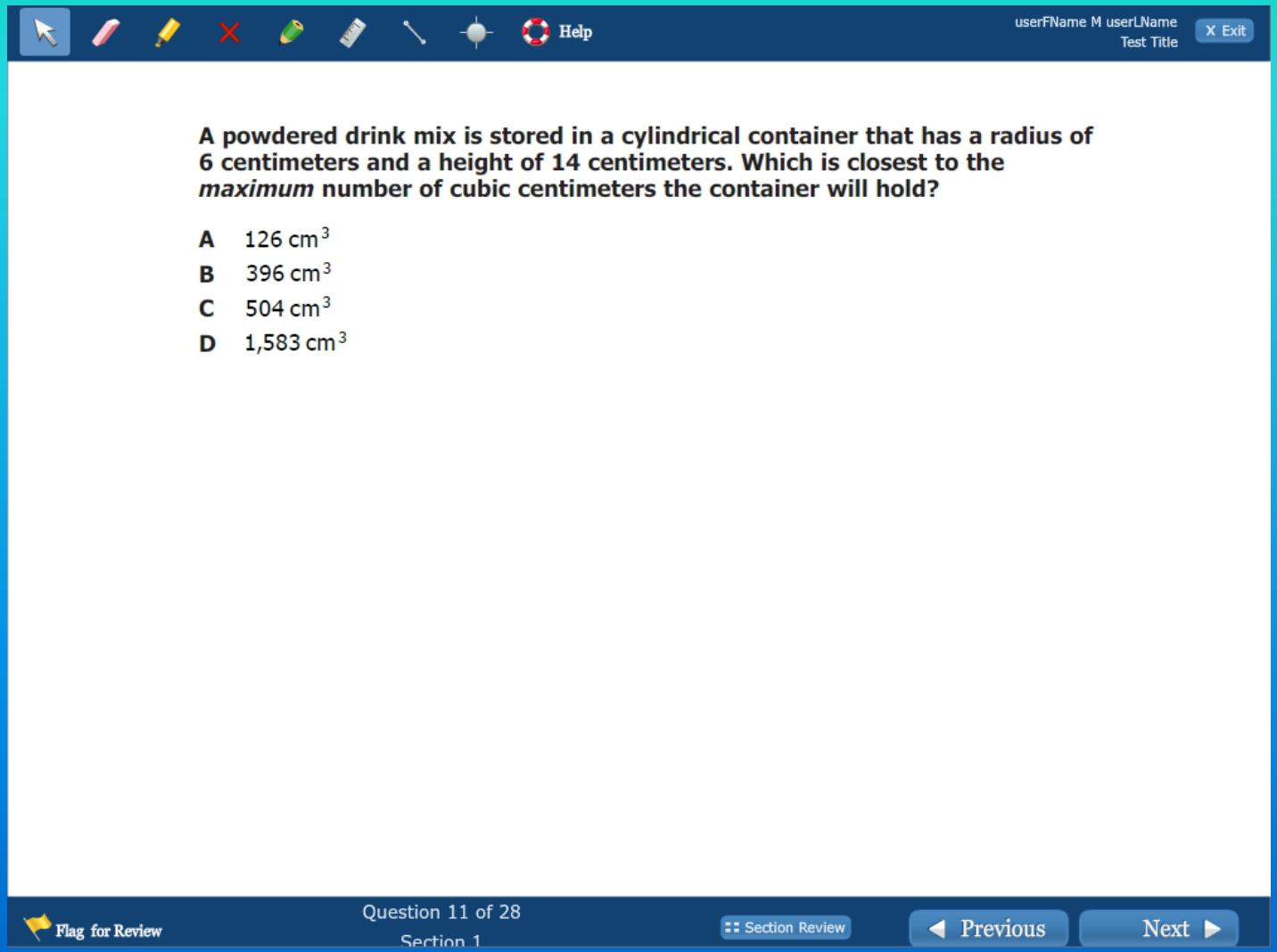
Question 10 of 28

Section Review

Previous

Next

Grade 7 Mathematics SOL Item



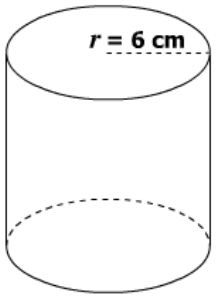
Grade 7 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit

A drink mix is stored in a cylindrical container that has a radius of 6 centimeters and a height of 14 centimeters.



r = 6 centimeters

h = 14 centimeters

$\pi \approx$ 3.14

$V =$ π · r^2 · h

Hint:
Click and drag the values of the radius, height, and π into the formula.

Which is closest to the volume of this container?

- **A** 264 cm^3
- **B** 528 cm^3
- **C** 1,583 cm^3

Flag for Review

Question 12 of 28

Section Review

Previous

Next

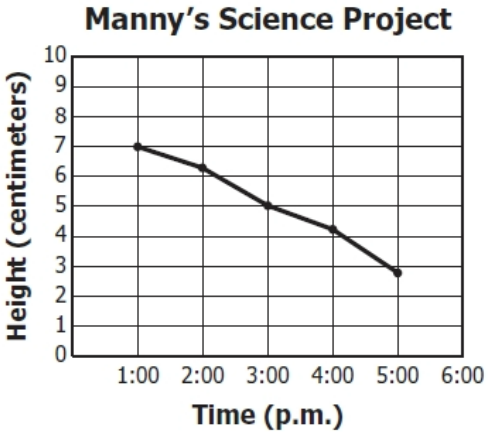
Grade 8 Mathematics SOL Item

Help

userFName M userLName
Test Title

X Exit

The line graph shows the heights of water in a container that Manny recorded over a 5-hour period for a science project.



If the height of the water continues to decrease as shown, which is closest to its height at 6:00 p.m.?

A 3.00 cm
B 2.75 cm
C 1.50 cm
D 0.25 cm

Flag for Review

Question 13 of 28
Section 1

Section Review

◀ Previous

Next ▶

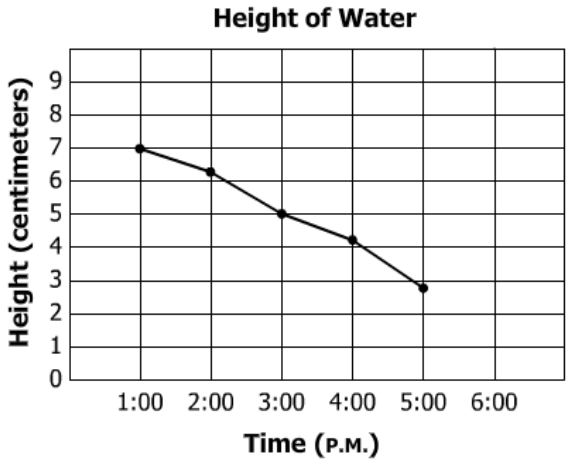
Grade 8 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit

The line graph shows the height of water in a container that Manny recorded over a 5-hour period.



Hint:
Place the pointer tool over each answer choice to extend the graph. Then select your answer.

If the height of the water continues to decrease as shown, which is closest to its height at 6:00 P.M.?

- **A** 3.25 cm
- **B** 2.75 cm
- **C** 1.50 cm

Flag for Review

Question 14 of 28

Section Review

Previous

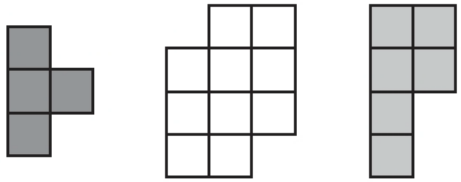
Next

Grade 8 Mathematics SOL Item

Help

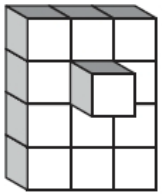
userFName M userLName
Test Title

X Exit

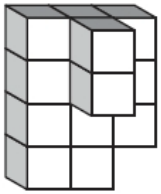


This shows 3 different views of a three-dimensional figure constructed from cubes. Which could be this figure?

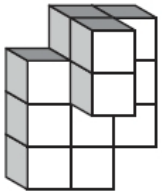
A



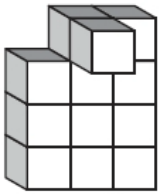
C



B



D



Flag for Review

Question 15 of 28

Section 1

Section Review

◀ Previous

Next ▶

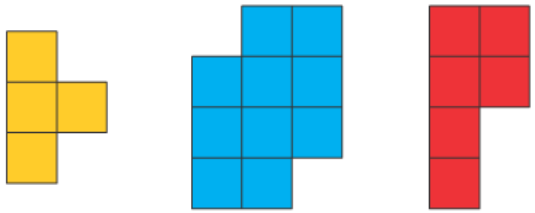
Grade 8 Mathematics VMAST Item

Help

userFName M userLName
Test Title

X Exit

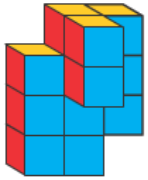
This shows 3 different views of a three-dimensional figure constructed from cubes.



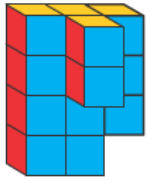
Top Front Left

Which could be this figure?

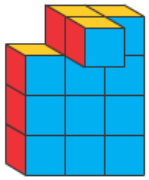
A



B



C



Flag for Review

Question 16 of 28

Section Review

Previous

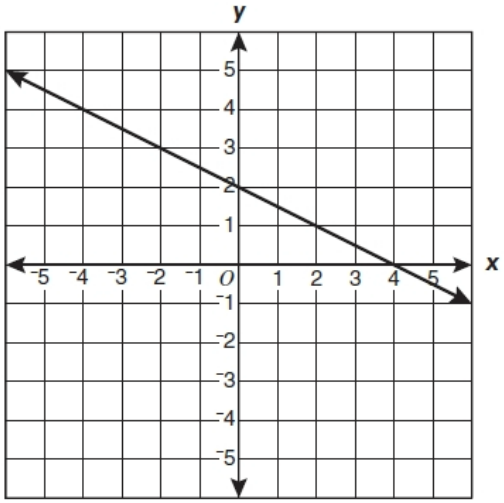
Next

End-of-Course Algebra I SOL Item

Help

userFName M userLName
Test Title

X Exit



Which could be an equation for the line shown on the grid?

A $y = -\frac{1}{2}x + 2$

B $y = -\frac{1}{2}x - 2$

C $y = -2x + 2$

D $y = 2x - 2$

Flag for Review

Question 17 of 28

Section 1

Section Review

◀ Previous

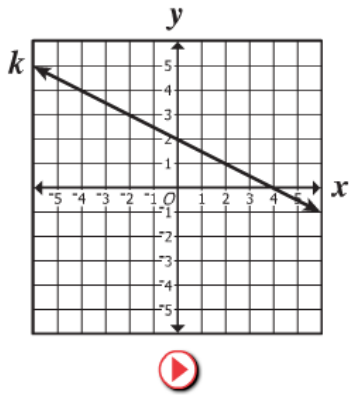
Next ▶

End-of-Course Algebra I VMAST Item

Help

userFName M userLName
Test Title

X Exit



Hint:

Click on to see the slope of the graphed line.

$$\text{Slope} = \frac{\text{change in } y}{\text{change in } x}$$

Which equation best represents line k ?

- **A** $y = \frac{-1}{2}x + 2$
- **B** $y = \frac{-1}{2}x - 2$
- **C** $y = -2x + 2$

Flag for Review

Question 18 of 28

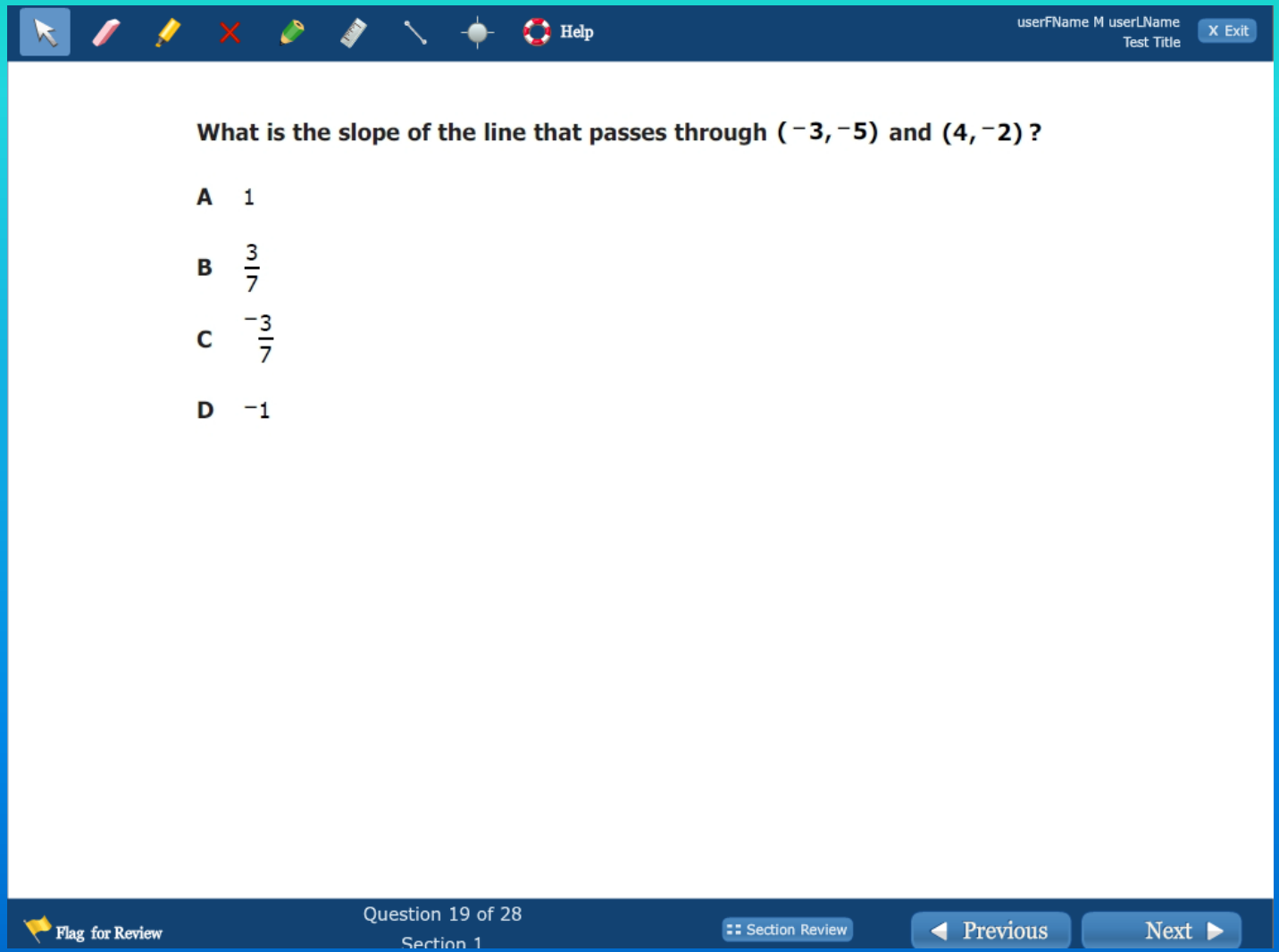
Section Review

Previous

Next

VIRGINIA DEPARTMENT OF EDUCATION

End-of-Course Algebra I SOL Item



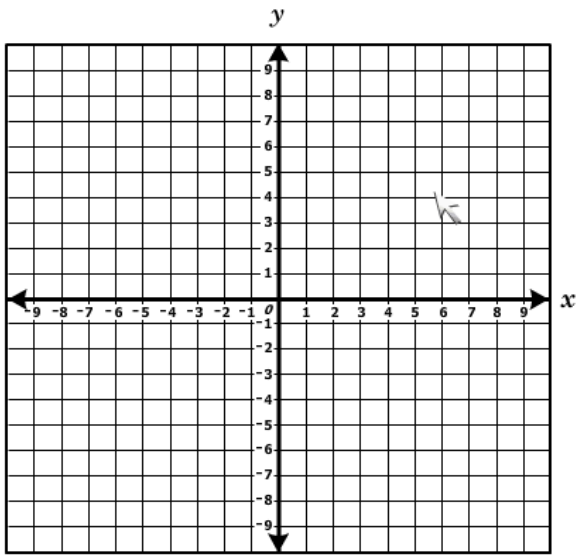
End-of-Course Algebra I
VMAST Item

Help

userFName M userLName
Test Title

X Exit

What is the slope of the line that passes through $(-3, -5)$ and $(4, -2)$?



Hint:

$\text{Slope} = \frac{\text{change in } y}{\text{change in } x}$

Click on the grid to plot points. A line will extend through the two points.

○ **A** $\frac{3}{7}$

○ **B** $\frac{-3}{7}$

○ **C** $-1\frac{7}{3}$

Flag for Review

Question 20 of 28

Section 1

Section Review

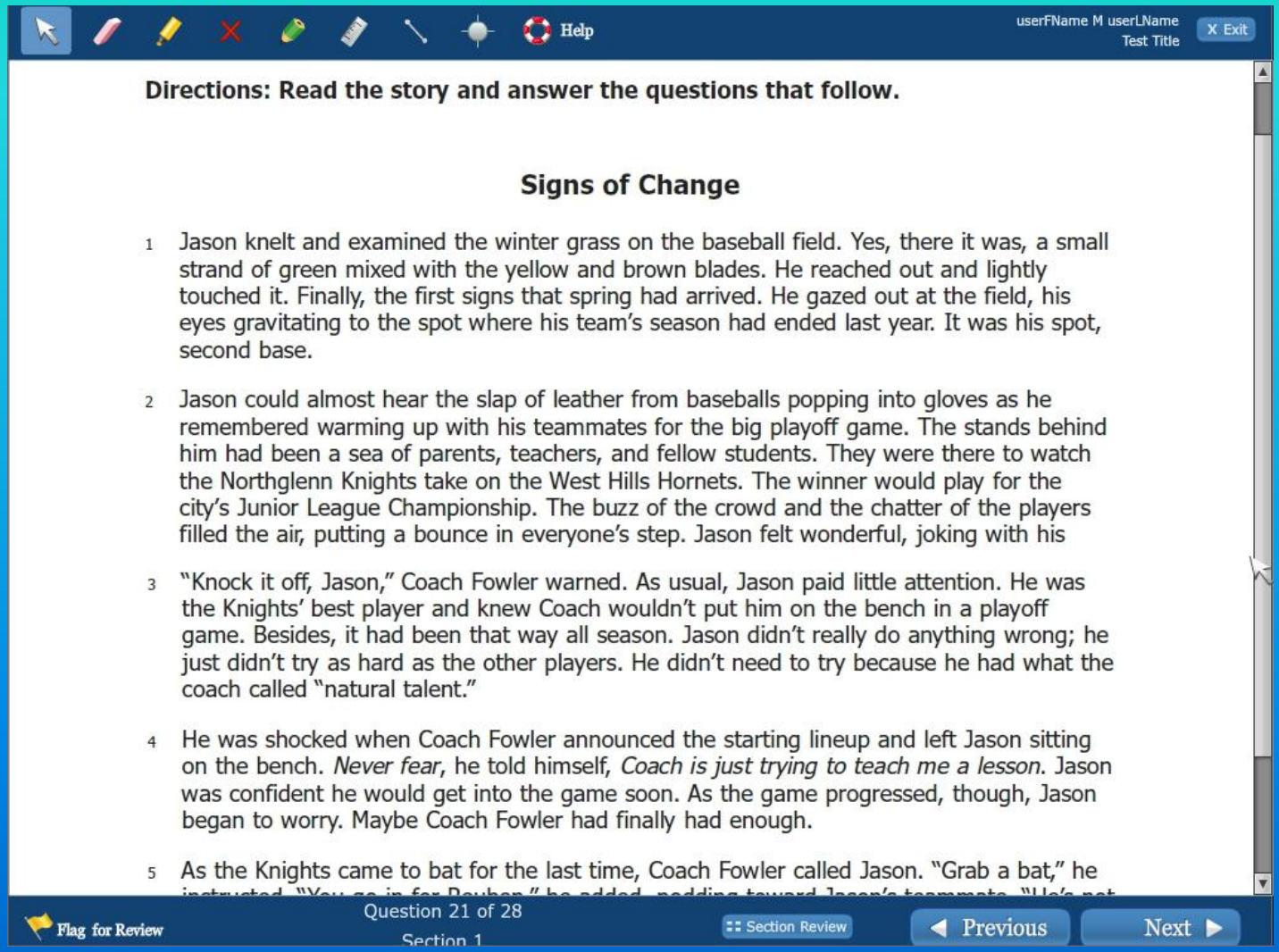
◀ Previous

Next ▶

Math Supports and Simplifications

- Provided hint boxes with formulas, strategies, and reminders
- Color coded important information
- Provided online manipulatives
- Simplified numbers
- Simplified language in stems
- Added or altered graphics
- Reduced answer options from 4 to 3

Grade 8 Reading SOL Passage



Grade 8 Reading
VMAST Passage

Help

userFName M userLName
Test Title

X Exit

Signs of Change

This passage contains a flashback, which is a technique where a part of a story describes something that happened in the past. In this story, Jason is remembering an event that happened during last year's baseball season.

1 Jason knelt and examined the winter grass on the baseball field. Yes, there it was, a small strand of green mixed with the yellow and brown blades. He reached out and lightly touched it. Finally, the first signs that spring had arrived. He gazed out at the field, his eyes gravitating to the spot where his team's season had ended last year. It was his spot, second base.

Page 1 of 4

Directions: Read the story and answer the questions that follow.

Flag for Review

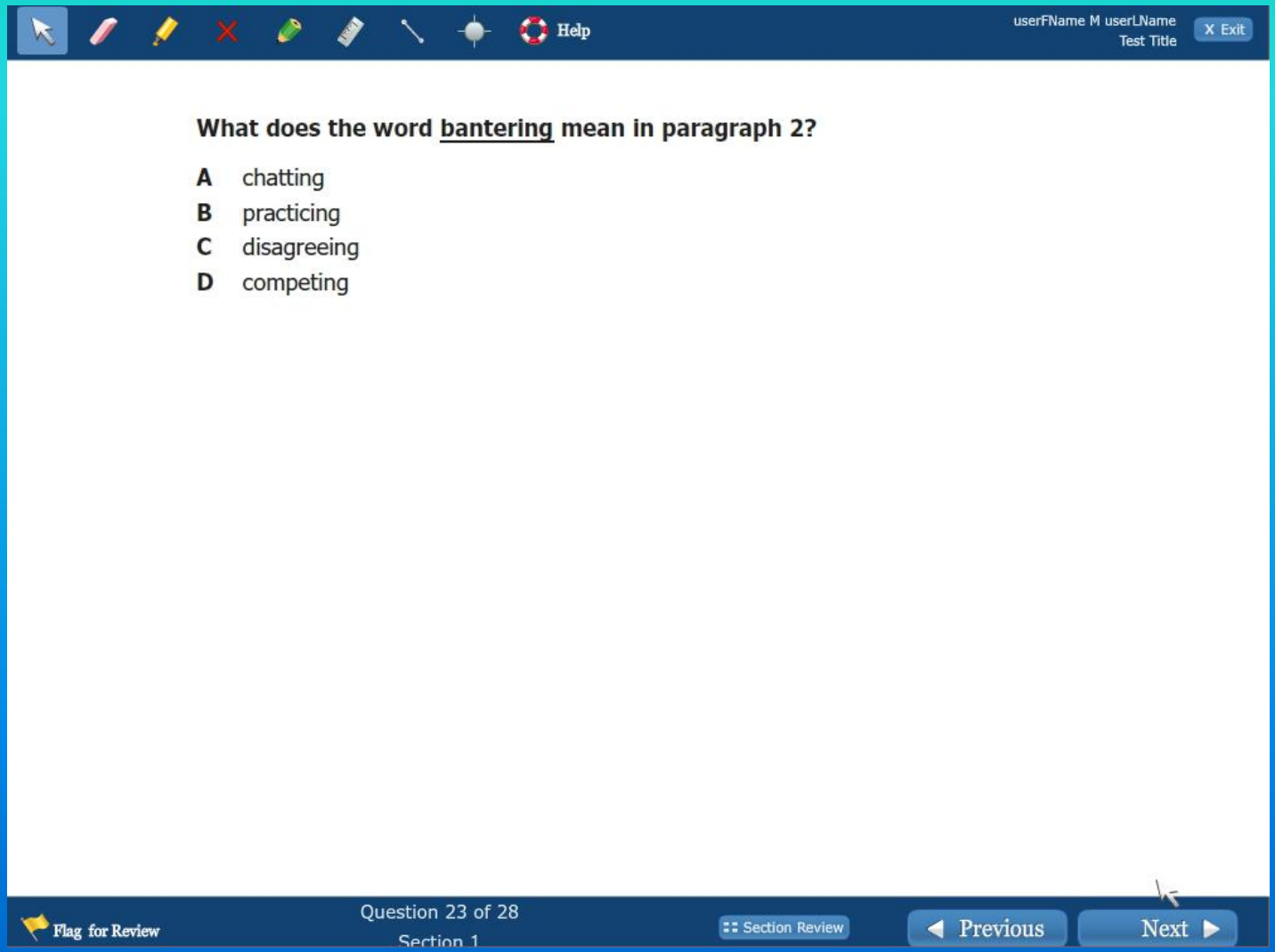
Question 22 of 28

Section Review

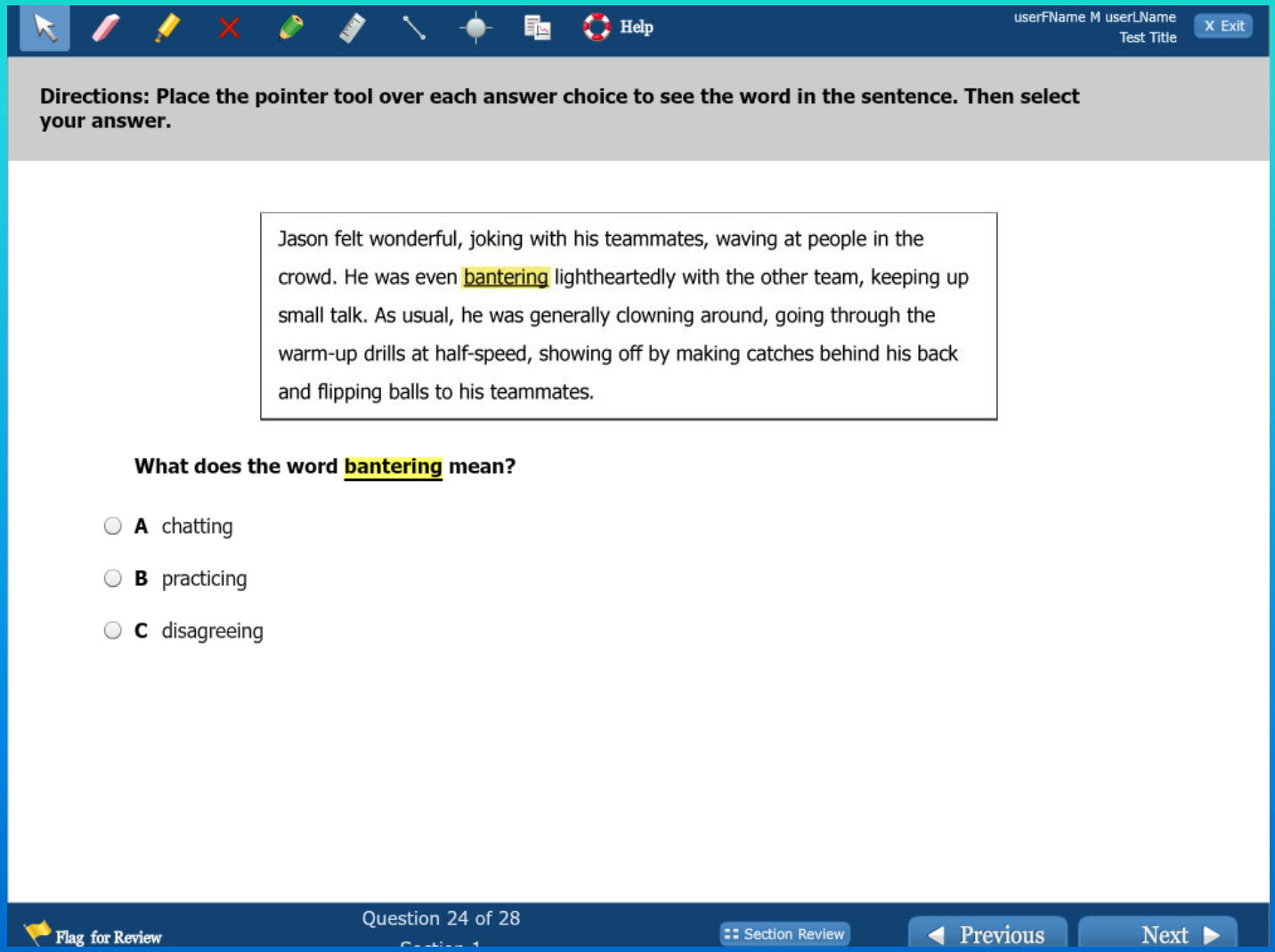
Previous

Next

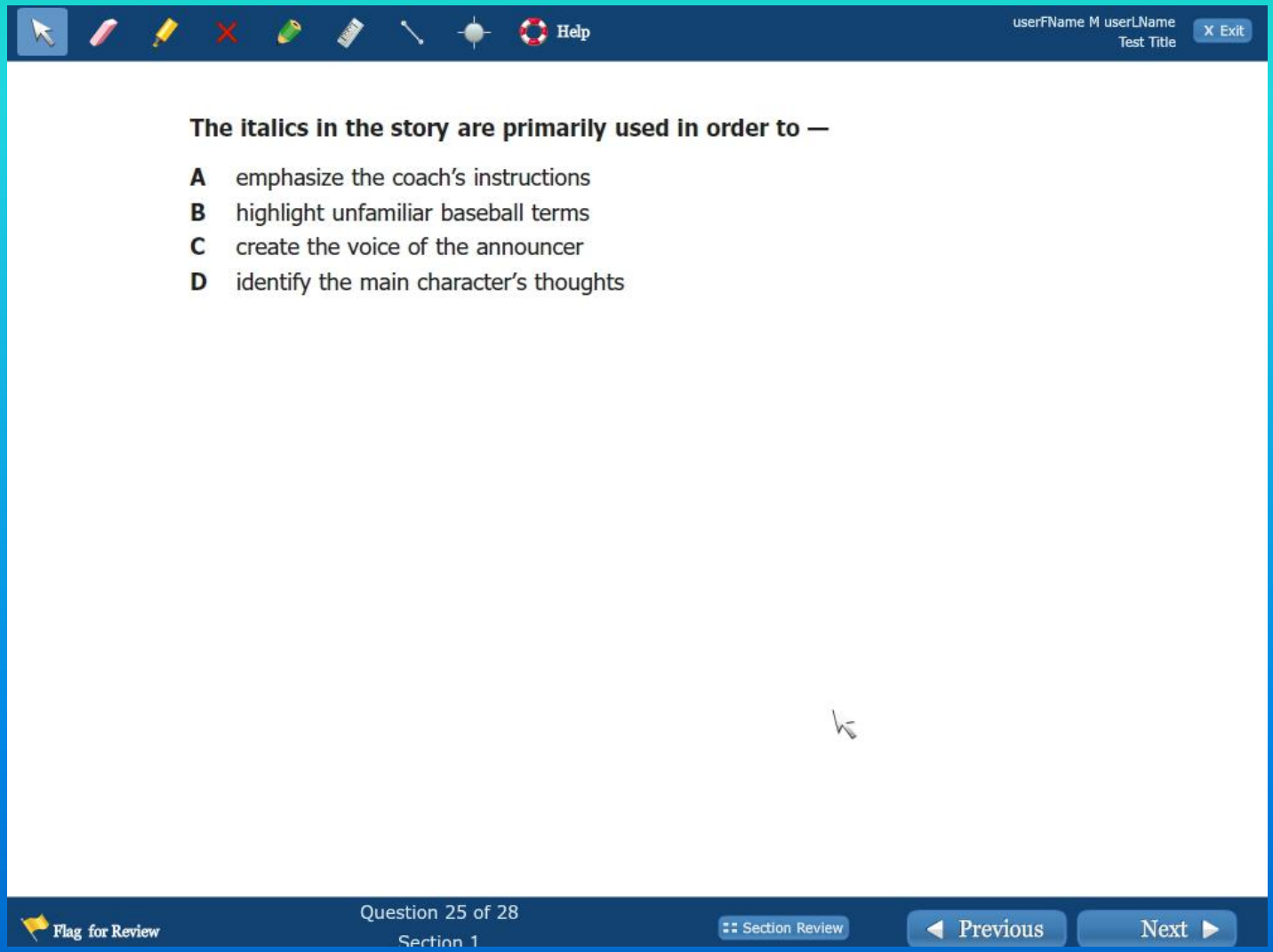
Grade 8 Reading SOL Item



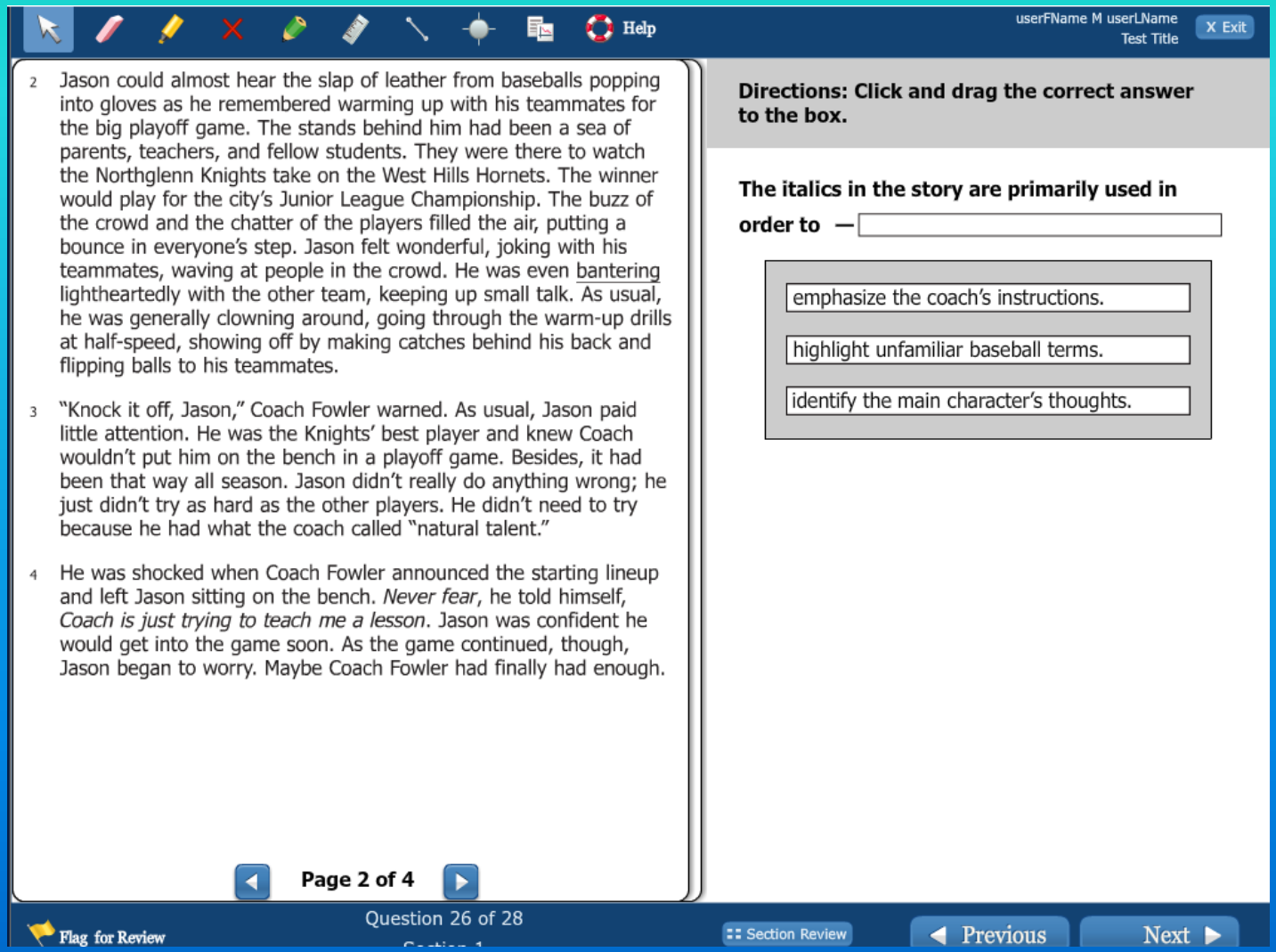
Grade 8 Reading
VMAST Item



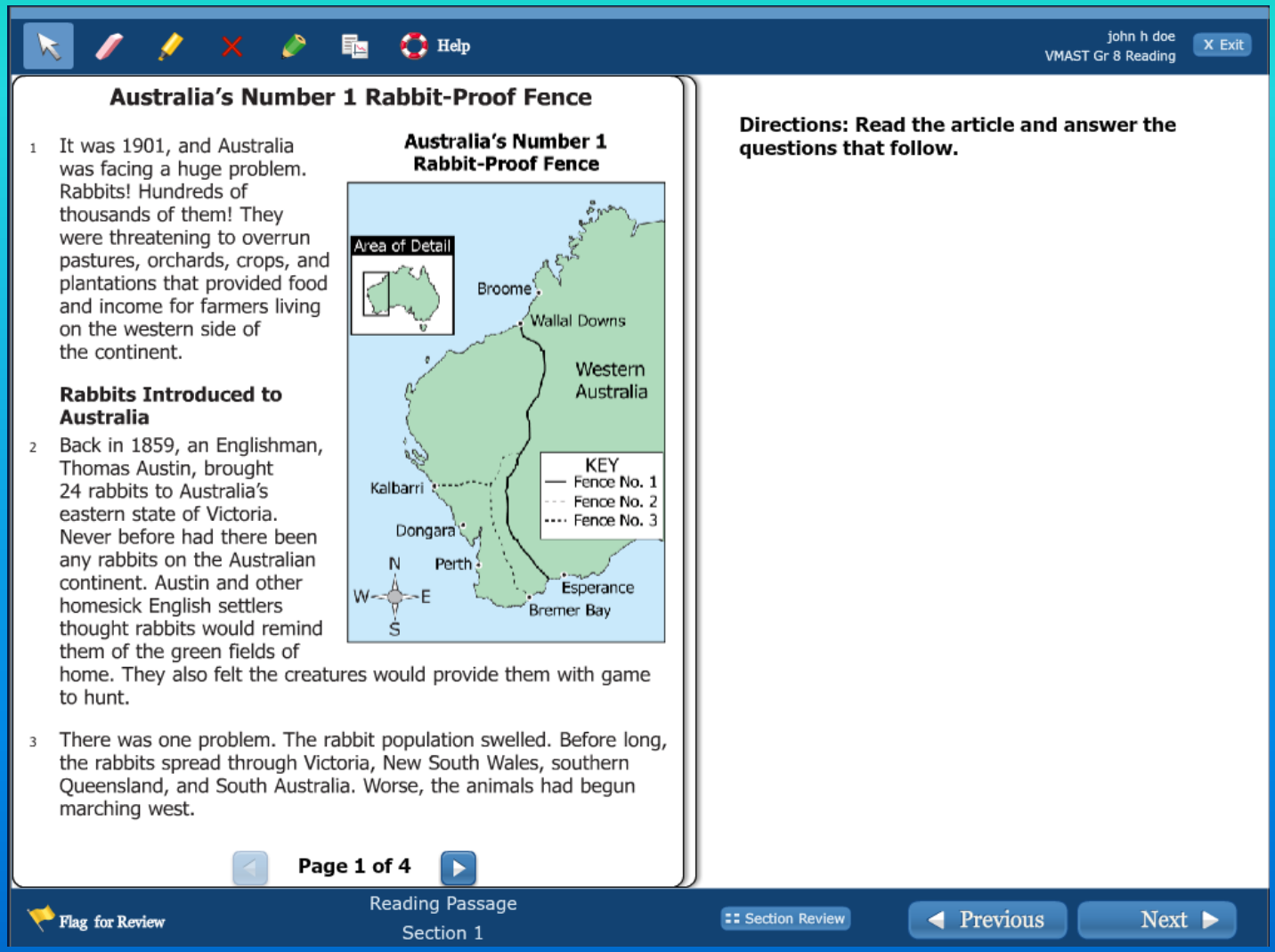
Grade 8 Reading SOL Item



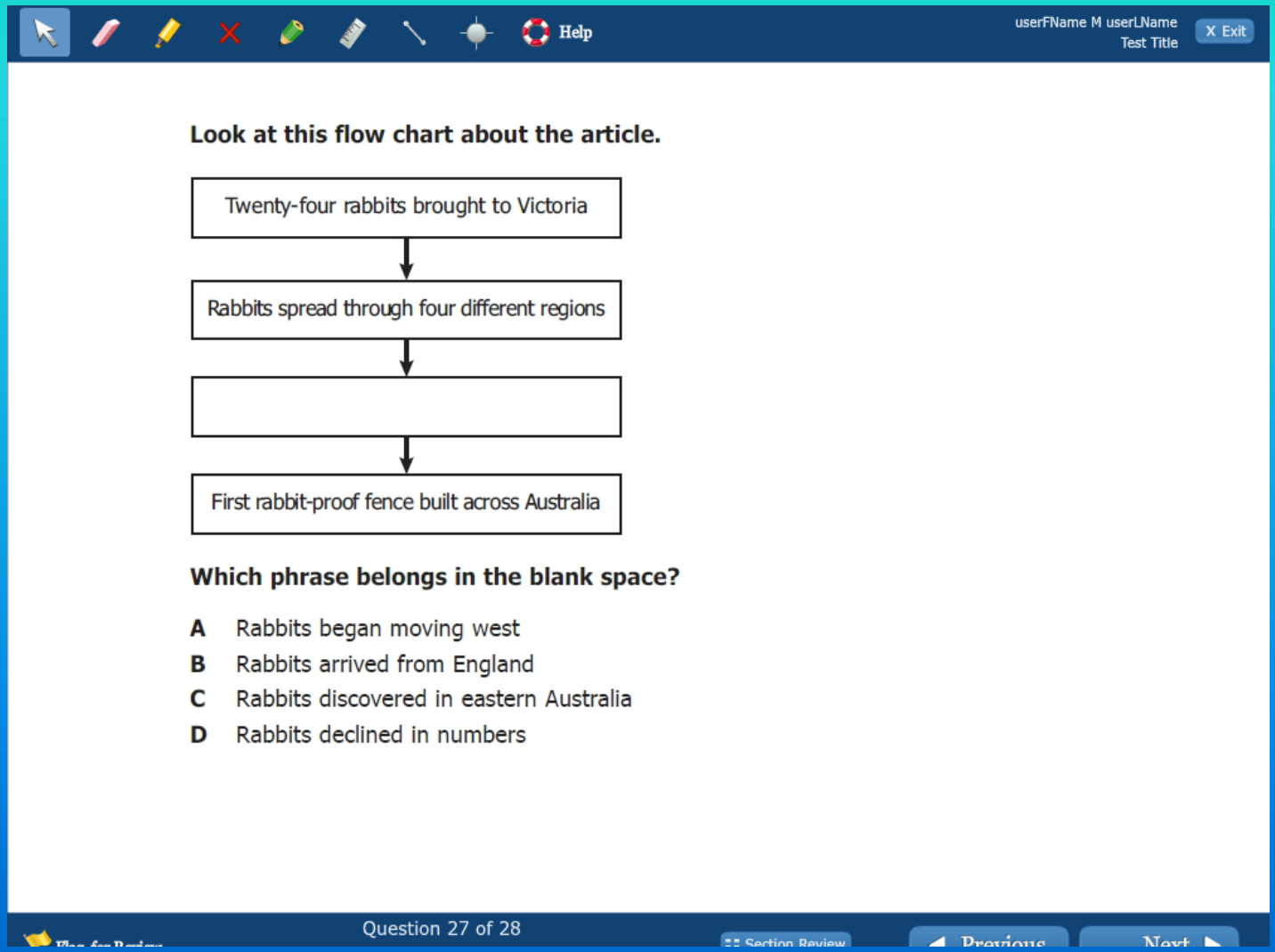
Grade 8 Reading
VMAST Item



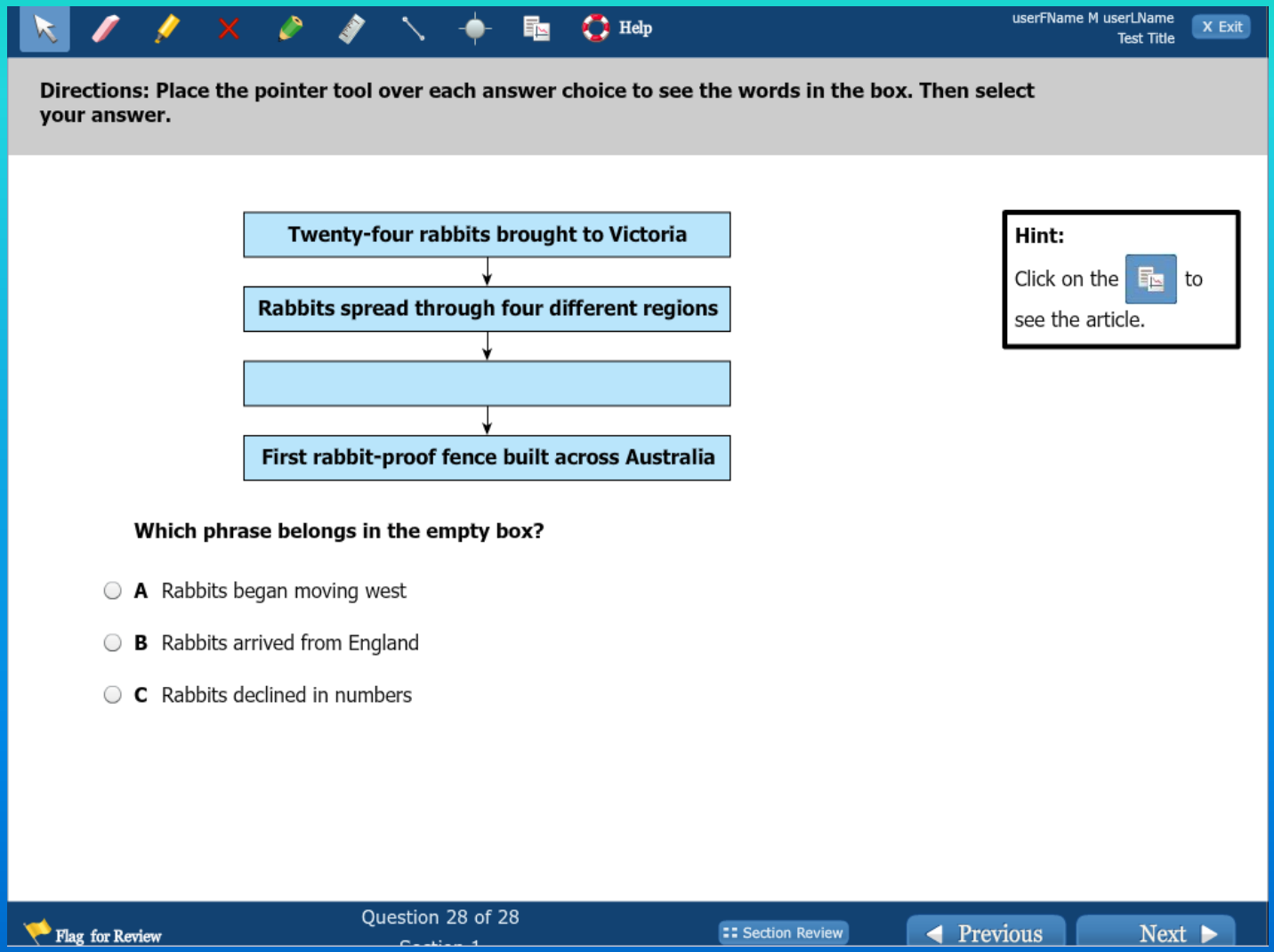
Grade 8 Reading
VMAST Passage



Grade 8 Reading SOL Item



Grade 8 Reading
VMAST Item



Reading Supports and Simplifications

- Shortened reading passages without changing reading level
- Excerpted relevant sections of reading and presented them with items
- Provided hint boxes to remind students to return to passage when item referred to passage as a whole

Reading Supports and Simplifications (cont.)

- Used graphic organizers to present passage information
- Simplified language in stems
- Highlighted key words
- Reduced answer options from 4 to 3

Lessons Learned

- Involve educators in identifying supports
- Be judicious in use of supports: “over-supporting” can be confusing to students
- Provide practice with items. VMAST practice items and guides for teachers may be found at:

  http://www.doe.virginia.gov/testing/alternative_assessments/vmast_va_mod_achievement_stds_test/practice_items/index.shtml

Favorites Tools Help
SLR_Partial_H Suggested Sites Google PEMSolutions SSWS VDOE Web Slice Gallery
Modified Achievement Standards Test...
gov Online Services | Commonwealth Sites | Help | Governor
RGINIA DEPARTMENT OF EDUCATION

VIRGINIA MODIFIED ACHIEVEMENT STANDARDS TEST (VMAST)

VMAST PRACTICE ITEMS

The VMAST practice items give students the opportunity to use research-based supports and simplifications that have been applied to existing online items to make them more accessible for students with disabilities. The practice items are samples only. They do not cover all content for the grade level nor do they provide examples of all item types or functionality that may be found in the field test items.

For technical assistance, see Technical Suggestions for Opening the Practice Items (PDF)

Mathematics

VMAST Mathematics Practice Items
Clicking on a link for the Practice Items will launch the items in a browser window.

Grade/Course	Practice Items	Practice Items – Audio Version	Practice Item Guides (PDF)
Grade 3	Practice Items	Practice Items – Audio	Guide
Grade 4	Practice Items	Practice Items – Audio	Guide
Grade 5	Practice Items	Practice Items – Audio	Guide
Grade 6	Practice Items	Practice Items – Audio	Guide
Grade 7	Practice Items	Practice Items – Audio	Guide
Grade 8	Practice Items	Practice Items – Audio	Guide
Algebra I	Practice Items	Practice Items – Audio	Guide

Application to Next Generation Assessments

Use principles of universal design in item development

- **Reduce language load**
- **Simplify presentation of items**

Maximize the use of technology to develop additional supports and simplifications

Contact Information:

Shelley Loving-Ryder
Virginia Department of Education

student_assessment@doe.virginia.gov
(804) 225-2102

Best Practices for Assessing Low Performing Students in NextGen, Grade Level Assessments

Ideas from an Assessment Design and Psychometric Perspective

Steve Ferrara

May 22, 2012

Overview

- Background and context
- Typical grade level test design
- How did programs go about designing and developing AA-MAS for low performing students?
- Ideas for NextGen grade level assessments
- What do we need to make these ideas work?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Background and context

- High school special ed teacher in Massachusetts: mild and moderate disabilities
- Work on alternate assessments in several states
- GSEG grant award: AA-MAS design research, development, and tryouts

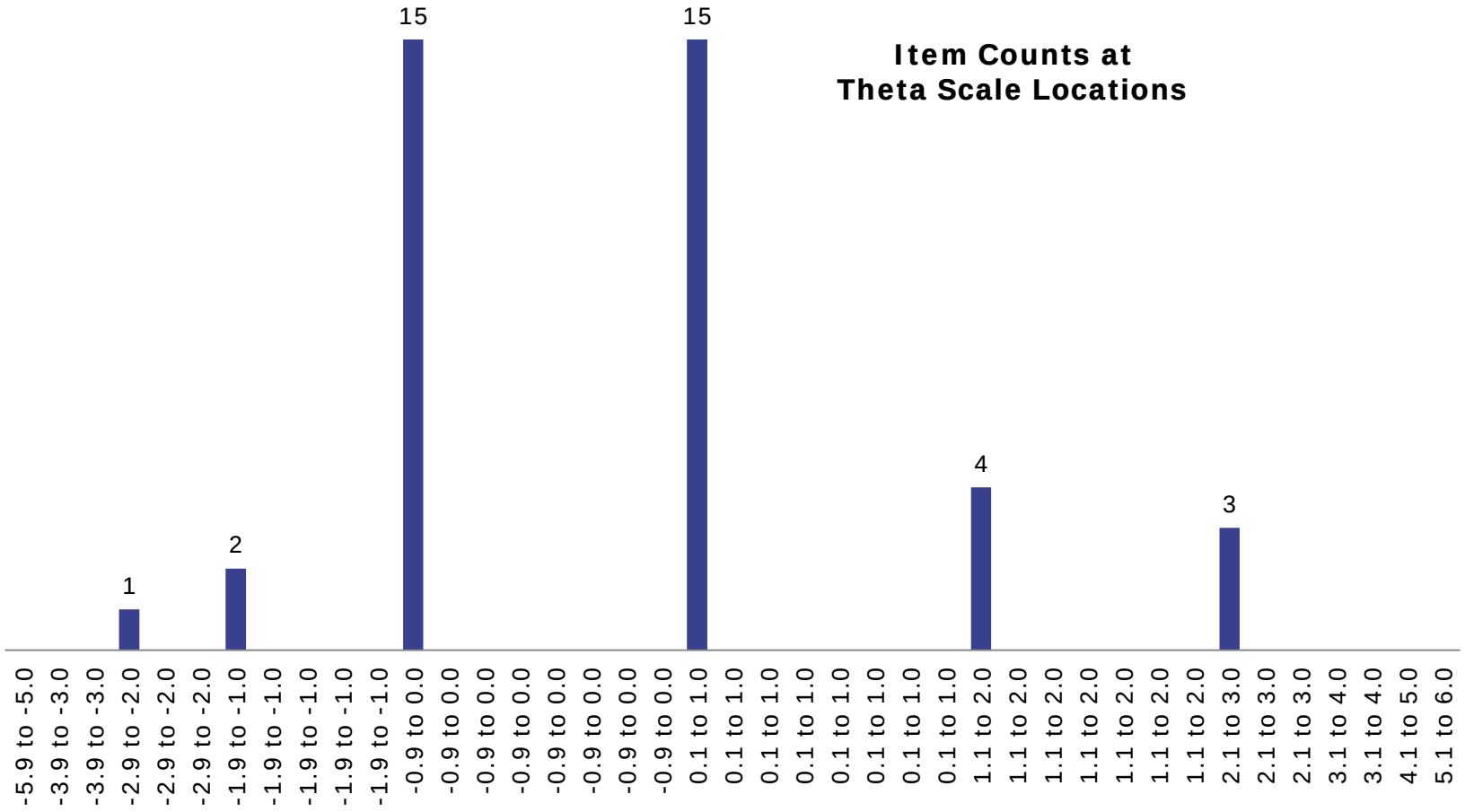
Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Typical grade level test design

- How we typically design grade level assessments
- What we typically get from them

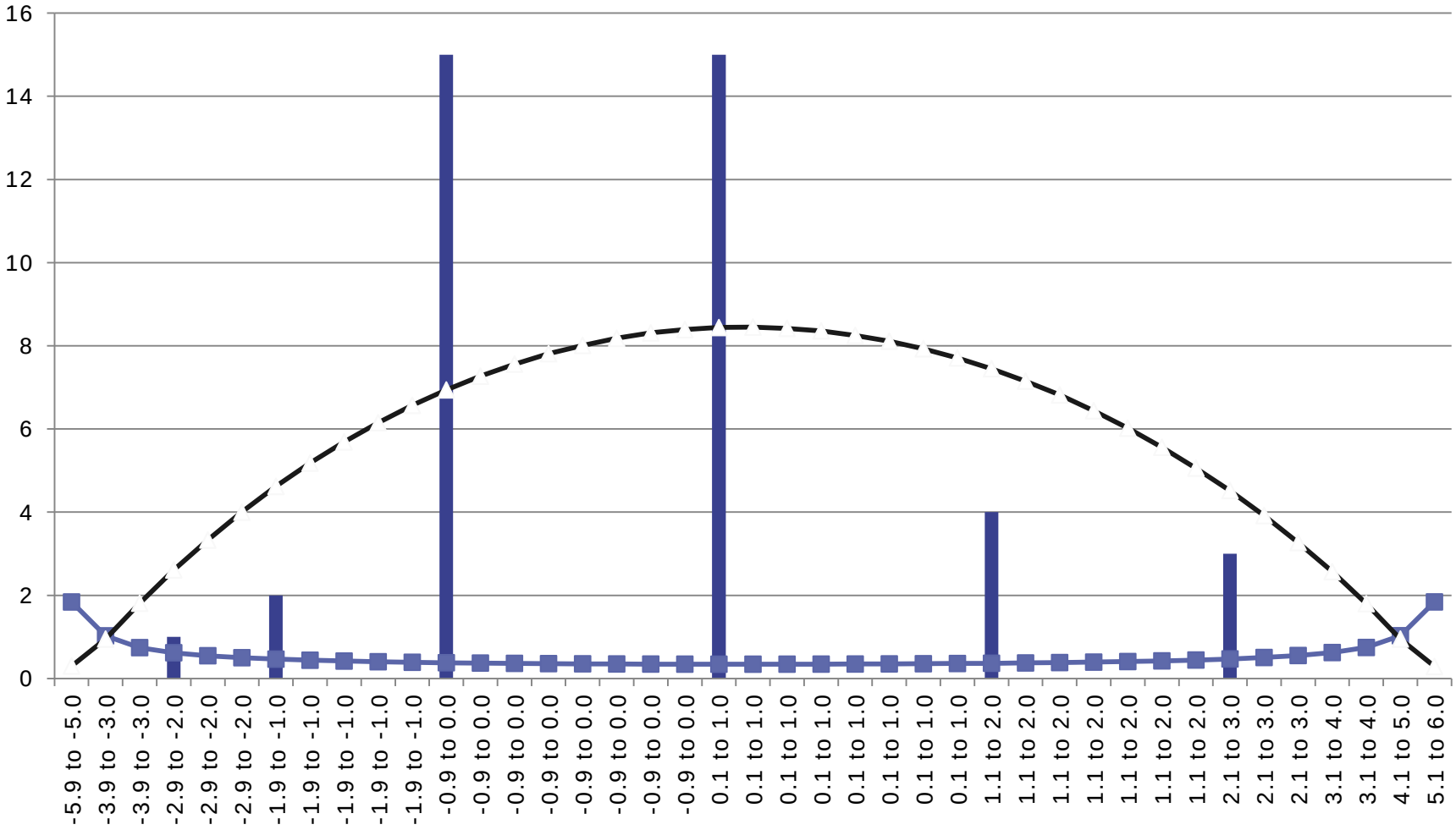
Copyright © 2011 Pearson Educatioideasn, Inc. or its affiliates. All rights reserved.

Real test, typical grade level test design



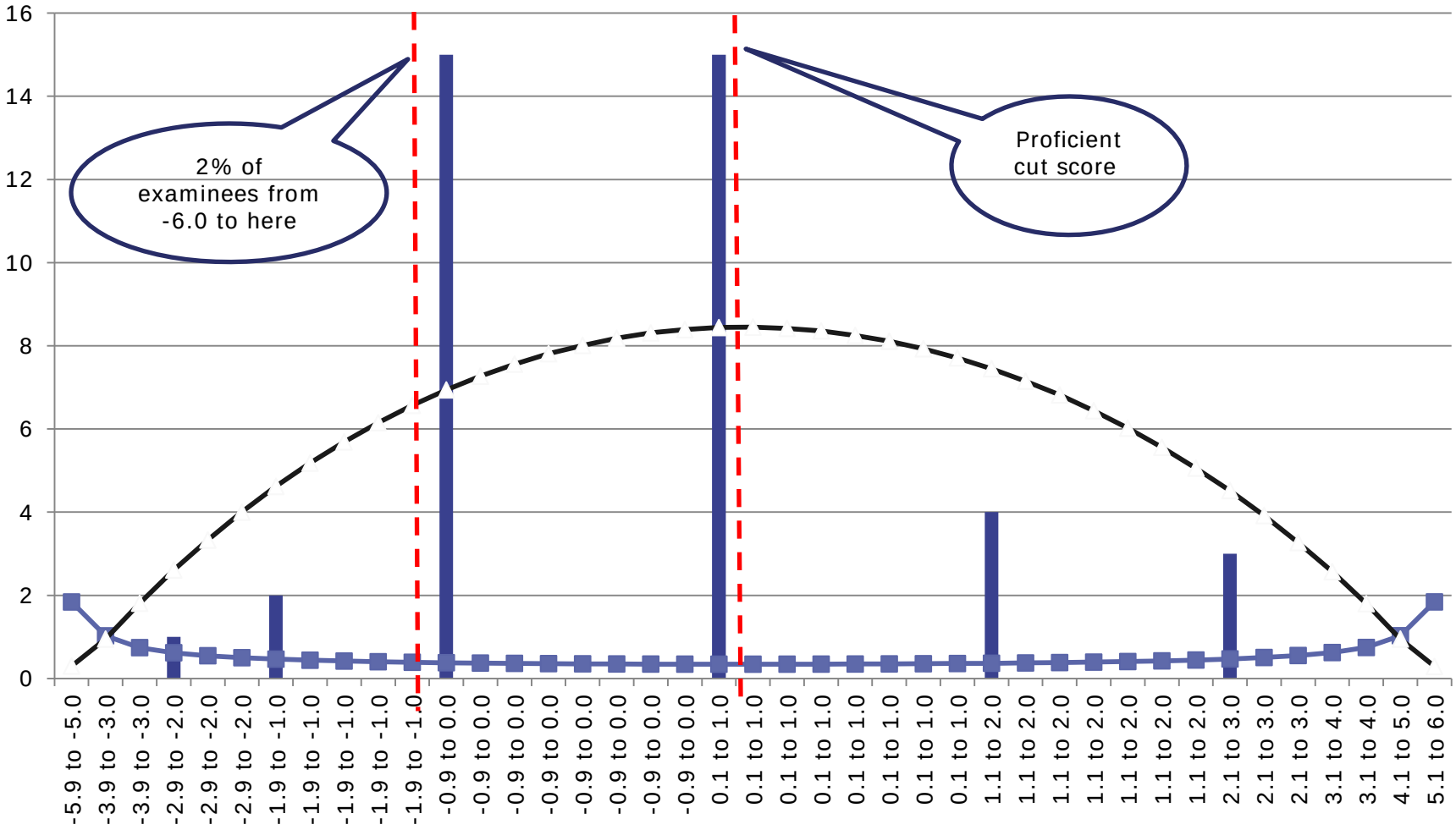
Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Each item adds psychometric information, mostly about nearby examinees



Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Item locations, score information, score standard errors



Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

So what do you see? What does this test give us?

- Lots precision for students in the middle of the scale, less precision for students at low and high ends of the proficiency scale
- α ~.85 and higher
- Result is good test targeting for most examinees, less so for low (and high) performing students

- Whether stated in assessment design and psychometric terms or not, a goal of AA-MAS was to target tests specifically for low performing students with disabilities
  - That is, to create more appropriately targeted tests of grade level content standards for these students

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

How did programs go about designing and developing AA-MAS for low performing students with disabilities?

- Tried to make tests easier
  - While targeting on-grade content standards
- Successful programs
- Lots of research that yielded important empirical contributions to designing accountability assessments for low performing students with disabilities
  - E.g., methods to reduce cognitive load: *Test Accessibility and Modification Inventory* (TAMI); see http://peabody.vanderbilt.edu/tami.xml

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Ideas for targeting NextGen grade level assessments

- Solution ideas to start thinking and discussion; not **the** solution
- Adaptive testing approaches
  - Item level CAT
  - Multistage, multilevel testing—with a twist

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Computer-adaptive testing (CAT)

- This is item level, adaptive test targeting
  - The process…
- Benefits
  - Effective test targeting ➔ precise test scores for examinees at all locations on the test scale
  - Efficiency: often possible to reach a specified level of score reliability with shorter tests
  - Effectiveness: Low performing students faced with few items that are too difficult for them
- Interesting finding:
  - Higher performing students say…
  - Lower performing students say…

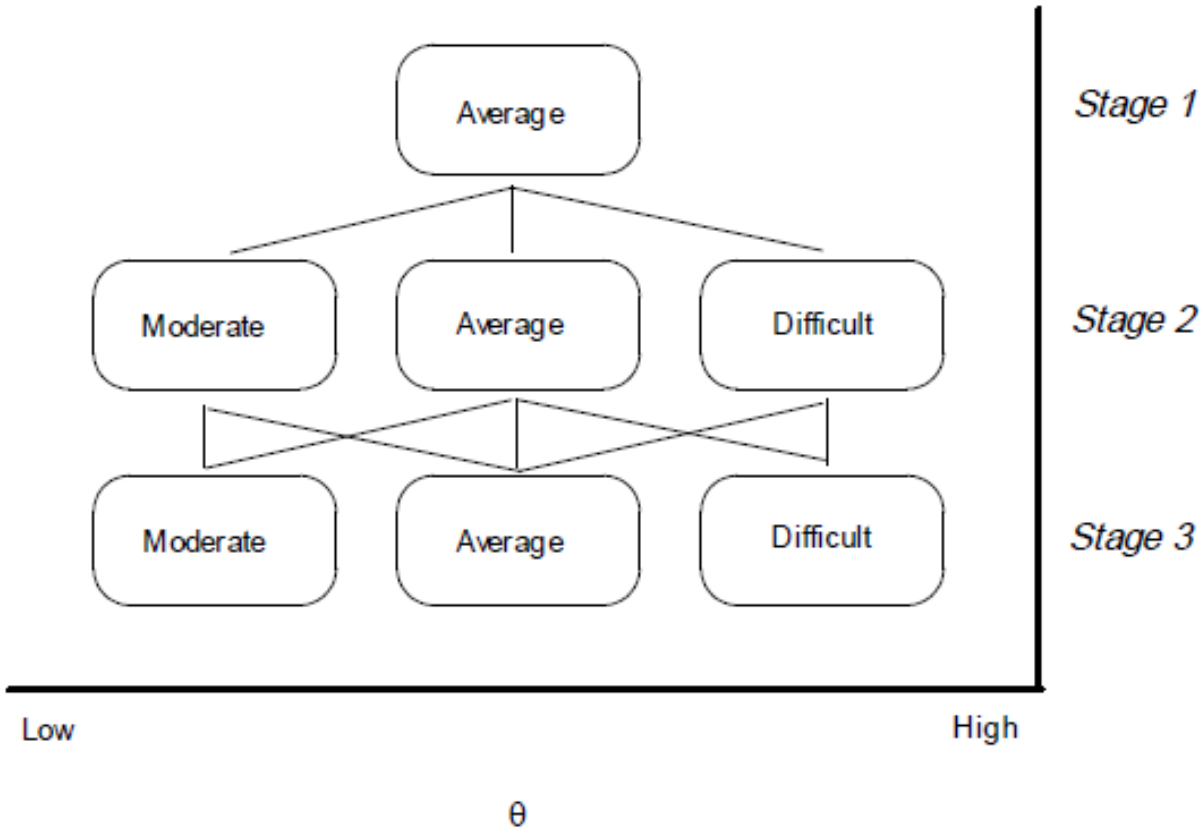
Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Multistage, multilevel testing

- Appears in many assessment design and measurement theory books (1970s and 1980s)
- Not in wide practice in educational testing, especially after item level CAT became operationally feasible in the mid-1980s
- Example
  - 15 item router test and three levels of 25 item tests, all on a single scale
  - High School Longitudinal Study of 2009 (HSLS:09); see http://nces.ed.gov/surveys/hsls09/

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

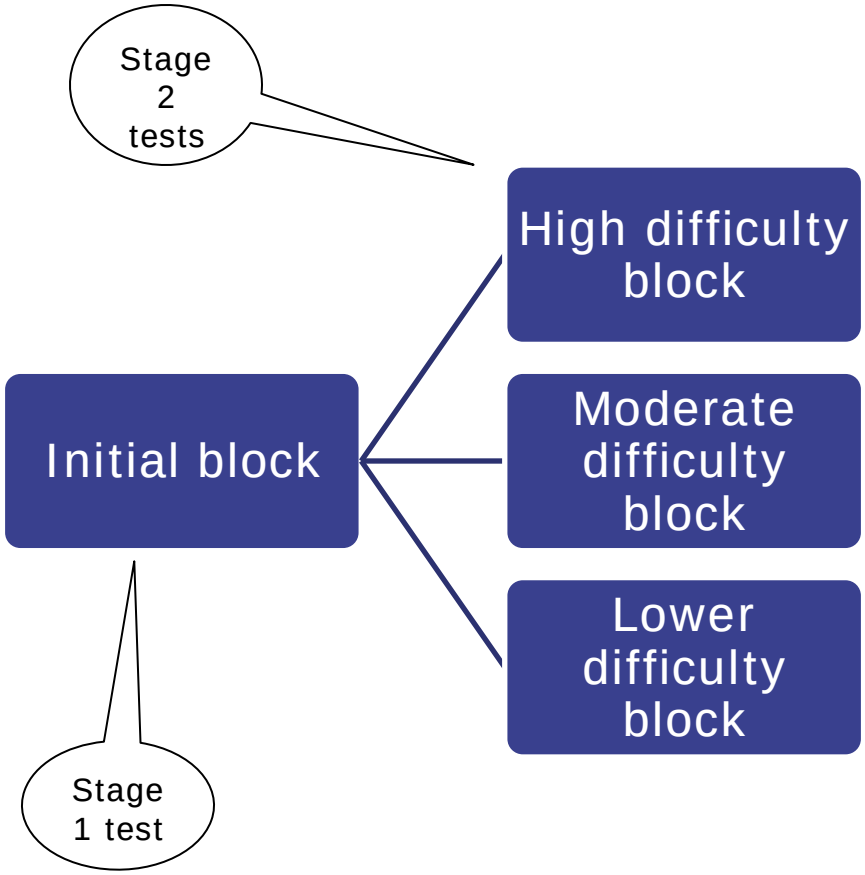
Schematic illustration



From Ariel (2005)

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

MSML design for NextGen grade level assessments



- All items and blocks on a single grade level scale
- All items and blocks selected from a single grade level pool
- Initial block representative of the full test blueprint
- Stage 1 score could be used for accountability reporting
- Stage 2 score:
  - Enhance the precision of the stage 1 score
  - Diagnostic targeting for low performing students with disabilities

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

What do we need for CAT and MSML tests? Adequate numbers of appropriately targeted items

- Need on-grade level, well aligned items, located at the lower end of proficiency scale
  - Must cover all content standards
- Often a challenge to produce enough of these for all content standards
  - Also, what about DOK levels? Why not other frameworks for addressing complexity?

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Generating enough items for appropriate targeting: research and experimentation

- Use easiest items on grade level tests to guide new item development
  - Clone them
  - Identify “easiness features” to guide additional item development
- Develop on grade, prerequisite cousins of current items
- Reduce the complexity of moderately difficult items: reduce cognitive load and build in easiness features
- Strive for simplicity rather than easiness
  - Focus on DOK levels 1 and 2…
  - Language simplicity, cognitive load, etc.
- Select on-grade reading texts with lower *Levels of Meaning* and *Knowledge Demands*, simpler *Structure*, and higher *Language Conventionality and Clarity* (see appendix A of the CCSS for ELA/Literacy)

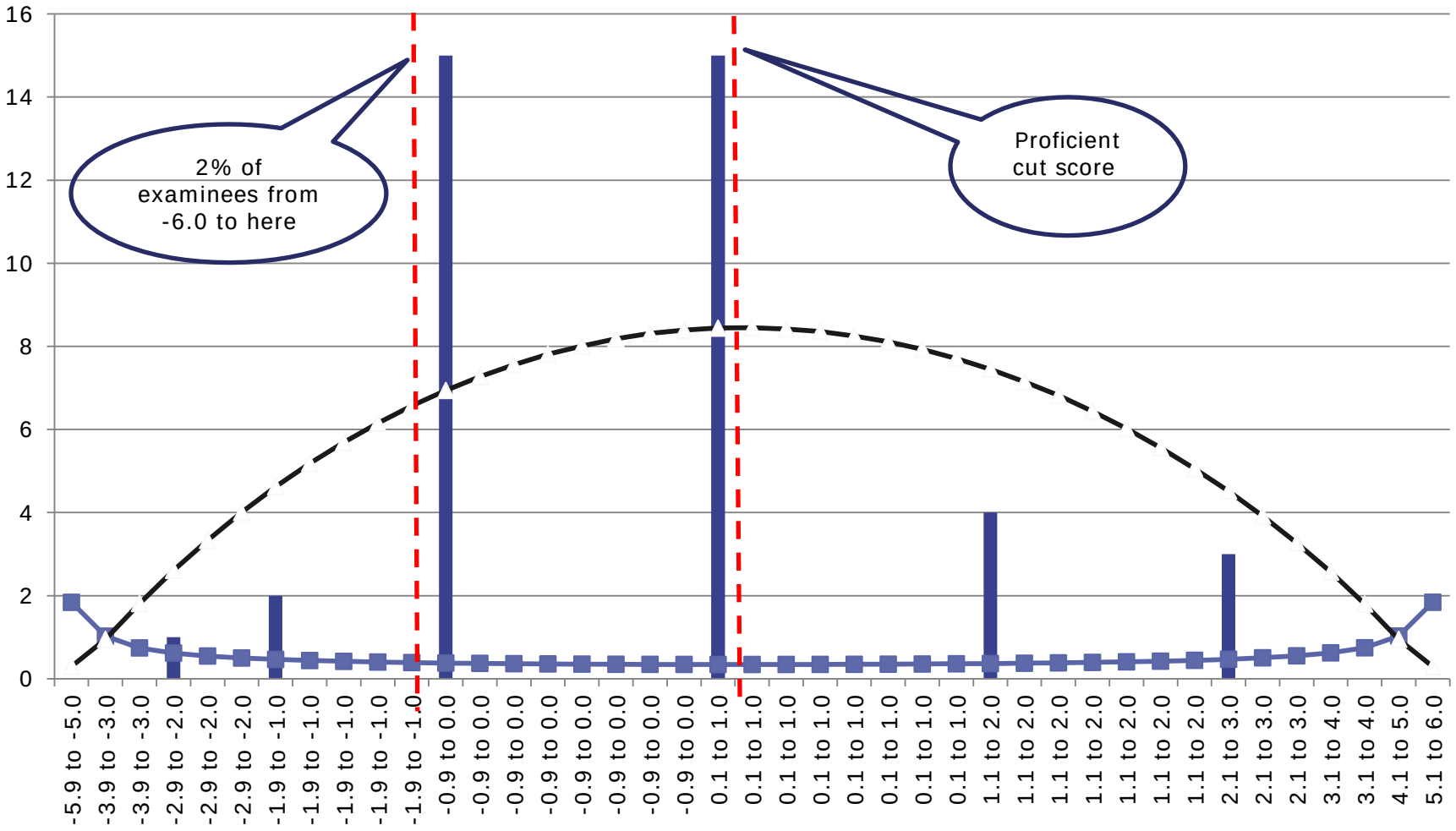
Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Closing comments (a)

- While grade level test scores are reliable overall, score reliability for lower performing students is not as good as the overall test score reliability suggests
  - Just take a look at the size of the standard errors for the lowest scores in your grade level program's scoring tables ➔
- Multistage, multilevel testing can work with fixed forms, administered online and on paper
  - Item pool size and replenishment requirements are not as demanding as for item level CAT
  - Not a widely known approach—possible reluctance, resistance
- These ideas—MSML design and targeting item development—require further conceptualization, design, vetting, research, and experimentation

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Reminder



Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Closing comments (b)

- Further conceptualization, design, and vetting
  - Would the policy community accept MSML testing? (cf. CAT)
  - Would state assessment programs accept it?
- Further research and experimentation
  - Ongoing, small scale item development projects focused on developing items that target grade level standards in the regions of score scales where many students with disabilities currently are performing
  - Build on the research from the last ~five years
  - Consider the empirical research on items with three response options: improved psychometric quality, minimal reduction of item difficulty—for all examinees
- Train item and task writers to hit difficulty targets more consistently (e.g., Ferrara et al., 2011)
  - We're not there yet

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

Thanks!

Steve Ferrara
Center for NextGen Learning and Assessment
steve.ferrara@pearson.com
612-581-6453

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.

References

Ariel, A. (2005). *Contributions to test-item bank design and management*. Doctoral dissertation, University of Twente, The Netherlands.

Ferrara, S., Svetina, D., Skucha, S., & Murphy, A. (2011). Test design with performance standards and achievement growth in mind. *Educational Measurement: Issues and Practice, 30* (4), 3-15.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24* (2), 3-13.

Copyright © 2011 Pearson Education, Inc. or its affiliates. All rights reserved.